

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 16:35:41 ; Search time 56.06 Seconds  
(without alignments)  
12.977 Million cell updates/sec

Title: US-09-786-648-3  
Perfect score: 62  
Sequence: 1 VEPGSHDSQ 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_0601.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62	100.0	12	21	Cholera toxin B/en
2	62	100.0	15	10	CTP3 epitope of th
3	62	100.0	15	16	Cholera toxin B an
4	62	100.0	21	21	Cholera toxin B/en
5	62	100.0	21	21	E. coli heat labil
6	62	100.0	23	16	Residues 50-64 of
7	62	100.0	26	4	Sequence of amino
8	62	100.0	41	6	Network polymer wh
9	62	100.0	46	6	Network polymer wh
10	62	100.0	47	4	Sequence of amino
11	62	100.0	93	16	ADP-ribosylating t

12	62	100.0	93	20	AAV41816
13	62	100.0	93	20	AAW95226
14	62	100.0	93	21	AAV68365
15	62	100.0	93	22	AAV68239
16	62	100.0	103	6	AAV50340
17	62	100.0	103	17	AAW04857
18	62	100.0	103	17	AAW94939
19	62	100.0	103	17	AAW06606
20	62	100.0	103	17	AAW06607
21	62	100.0	103	19	AAW08080
22	62	100.0	118	11	AAW04163
23	62	100.0	124	10	AAW93561
24	62	100.0	124	17	AAW06605
25	62	100.0	124	19	AAW59770
26	62	100.0	124	21	AAW96652
27	62	100.0	124	21	AAW98872
28	62	100.0	126	12	AAW12630
29	62	100.0	131	11	AAW04825
30	62	100.0	134	22	AAW73241
31	62	100.0	138	15	AAW50227
32	62	100.0	142	22	AAW73242
33	62	100.0	155	22	AAW73243
34	62	100.0	163	22	AAW73244
35	62	100.0	170	20	AAW94082
36	62	100.0	371	20	AAW01300
37	62	100.0	371	20	AAW67443
38	62	100.0	405	12	AAW11272
39	56	90.3	124	13	AAW28831
40	56	90.3	461	19	AAW74466
41	56	90.3	749	19	AAW80599
42	56	90.3	1338	19	AAW80600
43	49	79.0	461	20	AAW32094
44	47	75.8	15	21	AAW15525
45	47	75.8	15	21	AAW15526

## ALIGNMENTS

RESULT 1  
AAV87461  
ID AAV87461 standard; peptide; 12 AA.  
XX  
AC AAV87461;  
DT 03-JUL-2000 (first entry)  
XX  
DE Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:3.  
XX  
KW Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;  
KW beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;  
KW adjuvant; immune disorder; diarrhoea.  
XX  
OS Vibrio cholerae.  
OS Escherichia coli.  
XX  
PN WO200014114-A1.  
XX  
PD 16-MAR-2000.  
XX  
PF 07-SEP-1999; 99WO-GB02970.  
XX  
PR 07-SEP-1999; 98GB-0019484.  
XX  
PA (UYBR-) UNIV BRISTOL.  
XX  
PI Williams NA, Hirst TR;  
XX  
WPI; 2000-256943/22.  
XX  
PT Derivatives of Escherichia coli heat labile enterotoxins useful as  
PT immunomodulators and for treating diarrhea and which do not bind the  
PT glycolipid receptor GM-1 -

Escherichia coli V  
E. coli heat-labil  
Heat labile toxin  
E coli verotoxin-1  
Sequence of sub-un  
Synthetic cholera  
Heat labile entero  
Cholera toxin B su  
Cholera toxin B su  
Amino acid sequenc  
Cholera toxin B-su  
B subunit of the h  
Cholera toxin B su  
Amino acid sequenc  
Plant-optimized E.  
Plant-optimized V.  
GtB.1/CTB chimeri  
LTB-CTB fusion pro  
Recombinant exotox  
Sequence of LT-B-M  
Recombinant exotox  
Recombinant exotox  
LTB-CTP fusion pro  
Labile toxin (LT-B  
C. jejuni flagelli  
HSV-1 antigen/heat  
B subunit of CT  
Adhesin/V.cholerae  
Helicobacter pylor  
Adhesin/CTXA2B chi  
Cholera toxin B su  
Bovine rotavirus V

XX Disclosure; Page 15; 62pp; English.

XX The invention relates to peptide fragments of the *Escherichia coli* heat

CC labile enterotoxin (Etx) and its closely related homologue, cholera

CC toxin (Ctx) from *Vibrio cholerae* which do not bind to the ubiquitous

CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are

CC composed of one A subunit and five identical B subunits. The A subunit

CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-

CC ribosyltransferase activity, while the B subunits (CtxB and CtxB)

CC facilitate the entry of subunit A into the host cell via the binding and

CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible

CC for some of the effects of Etx and Ctx, it has been found that certain

CC effects of the toxins, such as immunomodulation, are not mediated

CC through GM-1 binding. The peptides of the invention are fragments of the

CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as

CC normal EtxB and CtxB subunits, except that they do not bind or cross link

CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.

CC They may also be used as an inhibitor for toxin-induced diarrhoea.

CC Therefore, the peptides may be used in the production of a composition

CC for treating, preventing and/or modulating a disease associated with an

CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463

CC represent preferred peptides of the invention, AAY87460 being

XX particularly preferred.

SQ Sequence 12 AA;

Query Match 100.0%; Score 62; DB 21; Length 12;

Best Local Similarity 100.0%; Pred. No. 3.2e-05;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSHIDSQ 12

DB 1 vevpsghidsq 12

RESULT 2

AAP93498

ID AAP93498 standard; protein; 15 AA.

XX

AC AAP93498;

XX

XX 03-MAY-1990 (first entry)

DT CTP3 epitope of the Cholera toxin B subunit.

DE

XX CTP3 epitope of Cholera toxin B subunit; flagellin fusion protein;

KW vaccine; immunotherapy; ds;

XX W08910967-A.

XX

XX 16-NOV-1989.

XX

XX 05-MAY-1989; 89WO-US01932.

XX

XX 05-MAY-1988; 88US-0190570.

XX

XX (PRAX-) PRAXIS BIOLOGICS INC.

XX (STRD) LEYLAND STANDFORD JUNIOR UNIV.

XX

XX Marjarian WR, Stocker BAD, Newton SMC;

XX

XX WPI: 1989-356496/48.

XX N-PSDB; AAN2414.

XX

XX New recombinant flagellin gene including sequence - for heterologous

PT epitope, and expressed fusion proteins, useful in vaccines and for prodn.

PT of antibodies.

XX

XX Disclosure; fig.4B; 137pp; English.

XX

XX This sequence corresponds to the CTP3 epitope of the Cholera toxin B

CC subunit. The DNA sequence encoding this ligates to othersynthetic

CC oligonucleotides to form a new recombinant gene. This encodes

CC a flagellin fusion protein which can be used in vaccines for immuno-

CC therapy.

SQ Sequence 15 AA;

Query Match 100.0%; Score 62; DB 10; Length 15;

Best Local Similarity 100.0%; Pred. No. 4.1e-05;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSHIDSQ 12

DB 1 vevpsghidsq 12

RESULT 3

AAR85125

ID AAR85125 standard; peptide; 15 AA.

XX

AC AAR85125;

XX

XX 13-JUN-1996 (first entry)

DT Cholera toxin B antigenic peptide fragment CTP3.

XX

XX Conjugate; cholera; B toxin; peptide fragment; microparticulate;

KW inert carrier; modified silica; thyroglobulin; oral vaccine;

KW immunisation; infection; insoluble; digestive tract; antigen;

KW intestines; antibodies; secretory; IgA class.

XX

XX *Vibrio cholerae*.

XX

XX W09529701-A1.

XX

XX 09-NOV-1995.

XX

XX 02-MAY-1995; 95WO-EP01661.

XX

XX 03-MAY-1994; 94IL-0109519.

XX

XX (YEDA) YEDA RES & DEV CO LTD.

XX

XX Marks RS, Mirelman D, Sela M;

XX

XX WPI: 1995-403805/51.

XX

XX Vaccines for oral immunisation against infecting agents, e.g.

PT cholera - comprise a conjugate of an antigen of an infecting agent

PT covalently bound to micro-particulate inert carrier, e.g. modified

PT aldehyde silica

XX

XX Claim 7; Page 25; 40pp; English.

XX

XX A compsn. comprising a conjugate of an antigenic cholera B toxin

CC peptide fragment (e.g. AAR85126-30 and esp. AAR85125), covalently

CC bound to a microparticulate inert carrier (e.g. modified silica or

CC thyroglobulin) can be used as an oral vaccine for immunisation

CC against cholera infection. The inert carrier is insoluble in the

CC digestive tract, allowing presentation of the antigen in the

CC intestines, where it will elicit antibodies mainly of the

CC secretory IgA class.

XX

XX Sequence 15 AA;

Query Match 100.0%; Score 62; DB 16; Length 15;

Best Local Similarity 100.0%; Pred. No. 4.1e-05;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSHIDSQ 12

DB 1 vevpsghidsq 12



ID AAR76748 standard; Protein; 23 AA.  
 XX  
 AC AAR76748;  
 XX  
 DT 18-MAR-1996 (first entry)  
 XX  
 DE Residues 50-64 of cholera toxin B subunit and FimH 224-226.  
 XX  
 KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;  
 KW FimA; FimF; FimG; receptor binding site; PCR; amplify; ss.  
 XX  
 OS Chimeric - Vibrio cholerae.  
 OS Chimeric - Escherichia coli.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1 /note= "Represents FimH residue 224"  
 FT Peptide 2..4 /note= "Linker peptide"  
 FT Peptide 5..19 /note= "Cholera toxin B subunit 50-64"  
 FT Peptide 20..22 /note= "Linker peptide"  
 FT Misc-difference 23 /note= "Represents FimH residue 226"  
 XX  
 PN W09520657-A1.  
 XX  
 PD 03-AUG-1995.  
 XX  
 PF 27-JAN-1995; 95WO-DK00042.  
 XX  
 PR 27-JAN-1994; 94US-0187166.  
 XX  
 XX (GXBI-) GX BIOSYSTEMS AS.  
 XX  
 PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;  
 DR WPI; 1995-275442/36.  
 DR N-PSDB; AAQ93061.  
 XX  
 PT Receptor specific bacterial adhesins - useful for targetting active  
 PT compounds and microbial cells to locations of receptors  
 XX  
 PS Example 3; Page 58; 152pp; English.  
 CC This sequence is encoded by a fragment of the the plasmid pLPA93  
 CC which was used in the production of fimb fusion genes comprising  
 CC the cholera toxin B subunit inserted into the fimb gene. This insert  
 CC shows the inclusion of the B subunit into the FimH protein at position  
 CC 224-226. The chimeric genes were then opt. further modified by insertion  
 CC of the hepatitis B virus surface antigen pre-S2 region into a different  
 CC position of the FimH adhesin of type 1 fimbriae. Restriction site handles  
 CC (BglII-sites) were introduced into the fimb gene, and the foreign  
 CC epitopes are then inserted in-frame. In the selected positions the  
 CC insertion of the epitopes did not significantly alter the adhesive  
 CC function of the FimH protein. The expression of the chimeric proteins  
 CC on the surface of fimbriae on bacterial hosts illustrated the possibility  
 CC of using bacterial adhesins as general presenters of foreign antigens and  
 CC epitopes. These chimeric genes may be used in the production of variant  
 CC FimH adhesins which may be useful for targetting active compounds  
 CC and microbial cells to locations comprising selected receptors to which  
 CC the adhesins bind.  
 XX  
 SQ Sequence 23 AA;  
 Query Match 100.0%; Score 62; DB 16; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 6.5e-05;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VEVPGSQHIDSQ 12  
 |||

Db 5 vevpgsqhidsq 16  
 RESULT 7  
 AAP30265  
 ID AAP30265 standard; Protein; 26 AA.  
 XX  
 AC AAP30265;  
 XX  
 DT 21-APR-1992 (first entry)  
 XX  
 DE Sequence of amino acids 50-75 of the cholera toxin B1 subunit which  
 DE carries an Arg at posns. 67 and 73.  
 XX  
 KW Cholera vaccine; therapy; E.coli infection; enterotoxin LT.  
 XX  
 OS Vibrio cholerae.  
 XX  
 PN EP95426-A.  
 XX  
 PD 30-NOV-1983.  
 XX  
 PF 26-MAY-1983; 83EP-0401052.  
 XX  
 PR 26-MAY-1982; 82PR-0009167.  
 XX  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX (INSP ) INST PASTEUR.  
 XX  
 PI Milhaud G, Raulais D, Rivaille P, Siffert O, Dodin A;  
 PI Guyongruaz A, Delmas A;  
 XX  
 DR WPI; 1983-834645/49.  
 XX  
 PT Cholera toxin B, sub-unit polypeptide(s) as vaccines and  
 PT medicaments - effective against Escherichia coli and Vibrio  
 PT cholerae infections, are prepd. by solid phase peptide synthesis  
 XX  
 PS Claim 7; Page 11; 13pp; French.  
 XX  
 CC The inventors claim cholera toxin B1 subunit sequences which carry  
 CC Arg in posn. 35, 67 or 73 (AAP30264, AAP30265, AAP30600). Tests have  
 CC shown that Arg in posns. 35, 67 and 73 play an important role in fixing  
 CC these toxins to cell walls. The peptides are used in the treatment of,  
 CC and vaccination against cholera infections and animal and human  
 CC infections due to E. coli (enterotoxin LT). The medicament may be  
 CC administered by oral, intraperitoneal, sub-cutaneous or intravenous  
 CC routes. For vaccines, pref. peptides having 15-30 AAs are injected  
 CC without carriers. Unit dose when used as a medicament is 50-500mg as  
 CC a vaccine 1-10mg of active cpd.  
 XX  
 SQ Sequence 26 AA;  
 Query Match 100.0%; Score 62; DB 4; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 7.4e-05;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VEVPGSQHIDSQ 12  
 |||  
 Db 1 vevpgsqhidsq 12  
 RESULT 8  
 AAP50439  
 ID AAP50439 standard; protein; 41 AA.  
 XX  
 AC AAP50439;  
 XX  
 DT 01-JAN-1980 (first entry)  
 XX  
 DE Network polymer which comprises a series of composite E. coli heat-  
 DE labile toxin (LT)/heat-stable toxin (ST) polypeptide repeating units.



XX Heat-labile enterotoxin; heat-stable toxin; vaccine; ss.  
 XX Synthetic.  
 XX WO8502611-A.  
 XX 20-JUN-1985.  
 XX 12-DEC-1984; 84WO-US02030.  
 XX 12-DEC-1983; 83US-0559469.  
 XX (SCRI-) SCRIPPS CLINIC RES.  
 XX Houghten RA;  
 XX WPI; 1985-159230/26.  
 XX New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -  
 XX useful for vaccination of mammals against the enterotoxin(s)  
 XX Claim 8; Page 100; 120pp; English.  
 XX The repeating units are bonded together by intramolecular  
 XX interpolypeptide cystine bonds formed between oxidized Cys residues  
 XX of the repeating units. This polypeptide may be used in the  
 XX vaccination of mammals for protection against the enterotoxins. The  
 XX composite polypeptide is made by solid phase synthesis or  
 XX recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,  
 XX X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.  
 XX Sequence 41 AA;  
 XX  
 XX Query Match 100.0%; Score 62; DB 6; Length 41;  
 XX Best Local Similarity 100.0%; Pred. No. 0.00012;  
 XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VEVPGSQHDSQ 12  
 DB 14 vevpgsqhdsq 25  
 RESULT 9  
 AAP50436  
 ID AAP50436 standard; protein; 46 AA.  
 XX AAP50436;  
 XX 01-JAN-1980 (first entry)  
 XX Network polymer which comprises a series of composite E. coli heat-  
 XX labile toxin (LT)/heat-stable toxin (ST) polypeptide repeating units.  
 XX Heat-labile enterotoxin; heat-stable toxin; vaccine; ss.  
 XX Synthetic.  
 XX WO8502611-A.  
 XX 20-JUN-1985.  
 XX 12-DEC-1984; 84WO-US02030.  
 XX 12-DEC-1983; 83US-0559469.  
 XX (SCRI-) SCRIPPS CLINIC RES.  
 XX Houghten RA;  
 XX WPI; 1985-159230/26.  
 XX

PT New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -  
 XX useful for vaccination of mammals against the enterotoxin(s)  
 XX Claim 8; Page 100; 120pp; English.  
 XX The repeating units are bonded together by intramolecular  
 XX interpolypeptide cystine bonds formed between oxidized Cys residues  
 XX of the repeating units. This polypeptide may be used in the  
 XX vaccination of mammals for protection against the enterotoxins. The  
 XX composite polypeptide is made by solid phase synthesis or  
 XX recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,  
 XX X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.  
 XX Sequence 46 AA;  
 XX  
 XX Query Match 100.0%; Score 62; DB 6; Length 46;  
 XX Best Local Similarity 100.0%; Pred. No. 0.00014;  
 XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VEVPGSQHDSQ 12  
 DB 16 vevpgsqhdsq 27  
 RESULT 10  
 AAP30600  
 ID AAP30600 standard; Protein; 47 AA.  
 XX AAP30600;  
 XX 21-APR-1992 (first entry)  
 XX Sequence of amino acids 350-75 of the cholera toxin B1 subunit which  
 XX carries an Arg at posns. 35, 67 and 73.  
 XX Cholera vaccine; therapy; E.coli infection; enterotoxin LT.  
 XX Vibrio cholerae.  
 XX EP95426-A.  
 XX 30-NOV-1983.  
 XX 26-MAY-1983; 83EP-0401052.  
 XX 26-MAY-1982; 82EP-0009167.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX (INSP ) INST PASTEUR.  
 XX Milhaud G, Raulais D, Rivaille P, Siffert O, Dodin A;  
 XX Guyongruaz A, Delmas A;  
 XX WPI; 1983-834645/49.  
 XX Cholera toxin B, sub-unit polypeptide(s) as vaccines and  
 XX medicaments - effective against Escherichia coli and Vibrio  
 XX cholerae infections, are prepd. by solid phase peptide synthesis  
 XX Claim 8; Page 11; 13pp; French.  
 XX The inventors claim cholera toxin B1 subunit sequences which carry  
 XX Arg in posn. 35, 67 or 73 (AAP30264, AAP30265, AAP30600). Tests have  
 XX shown that Arg in posns. 35, 67 and 73 play an important role in fixing  
 XX these toxins to cell walls. The peptides are used in the treatment of,  
 XX and vaccination against cholera infections and animal and human  
 XX infections due to E. coli (enterotoxin LT). The medicament may be  
 XX administered by oral, intraperitoneal, sub-cutaneous or intravenous  
 XX routes. For vaccines, pref. peptides having 15-30 AAs are injected  
 XX without carriers. Unit dose when used as a medicament is 50-500mg as  
 XX a vaccine 1-10mg of active cpd.  
 XX

SQL Sequence 47 AA;

Query Match 100.0%; Score 62; DB 4; Length 47;  
Best Local Similarity 100.0%; Pred. No. 0.00014;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
Db 22 vevpgsqhdsq 33

RESULT 11

AAR72545  
ID AAR72545 standard; peptide; 93 AA.

XX AC AAR72545;

XX DT 28-NOV-1995 (first entry)

XX DE ADP-ribosylating toxin (verotoxin-1 B-subunit).

XX KW ADP-ribosylating toxin; pertussis holotoxin; B-subunit;  
active site; E. coli heat labile toxin; verotoxin-1;  
Bordetella pertussis vaccines.

XX OS Bacteria sp.

XX XX EP646599-A.

XX PD 05-APR-1995.

XX XX 23-AUG-1994; 94EP-0306219.

XX PR 24-AUG-1993; 93US-0110947.

XX PR 31-MAY-1994; 94US-0251121.

XX PA (CONN-) CONNAUGHT LAB LTD.

XX XX (UYAL-) UNIV ALBERTA.

XX PI Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;  
PI Oomen R, Read RJ, Stein PE;

XX XX WPI: 1995-132623/18.

XX PT New modified forms of pertussis holotoxin - developed using  
PT crystalline forms of pertussis holotoxin and its complexes with  
PT other molecules

XX PS Disclosure; Fig 5; 54pp; English.

XX CC AAR72540-872545 are structurally equivalent B-subunits from three  
CC ADP-ribosylating toxins, pertussis holotoxin (PT), E. coli heat  
CC labile toxin (LT), and verotoxin-1 (VT). The structural  
CC information obtd. from these comparisons was used to identify  
CC sites which contribute to PT's biological activity. By modifying  
CC these sites the claimed PT mutants of the invention were produced.  
CC they can be used in the development of vaccines against Bordetella  
CC pertussis infection.

XX SQL Sequence 93 AA;

Query Match 100.0%; Score 62; DB 16; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.0003;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
Db 40 vevpgsqhdsq 51

RESULT 12

AAV41816

XX ID AAY41816 standard; peptide; 93 AA.

XX AC AAY41816;

XX DT 08-DEC-1999 (first entry)

XX DE Escherichia coli verotoxin-1 B-subunit.

XX KW ADP-ribosylating toxin; PT; crystalline pertussis holotoxin; ETA; DT;  
three-dimensional structure; LT; immunoprotective; infection.

XX OS Escherichia coli.

XX PN US5965385-A.

XX PD 12-OCT-1999.

XX XX 06-JUN-1995; 95US-0467974.

XX PR 22-AUG-1994; 94US-0292968.

XX PR 24-AUG-1993; 93US-0110947.

XX PR 31-MAY-1994; 94US-0251121.

XX PA (CONN-) CONNAUGHT LAB LTD.

XX PA (UYAL-) UNIV ALBERTA.

XX PI Read RJ, Stein PE, Cockle SA, Loosmore S, Klein MH, Armstrong GD;

XX PI Hazes B, Oomen RP;

XX XX WPI: 1999-579908/49.

XX DR New method for producing modified pertussis holotoxin -

XX PT Example 3; Fig 5; 41pp; English.

XX CC A method has been developed of producing a modified pertussis holotoxin,  
CC involving analysis of the 3-dimensional form of the crystalline  
CC holotoxin. The pertussis holotoxin modification process comprises:  
CC (1) identification of at least one amino acid (aa) residue of the  
CC holotoxin for modification by analysing the 3-dimensional form of the  
CC crystalline holotoxin, in relation to known information of the protein  
CC structure and function; (2) effecting mutagenesis (by removing or  
CC replacing a nucleotide sequence encoding at least one (aa) of a tox  
CC operon; and (3) expressing mutant tox box in a Bordetella organism to  
CC produce the modified holotoxin. This method is used for modifying  
CC pertussis holotoxin, by studying its 3-dimensional crystalline  
CC structure. Modifying the holotoxin, alters its biological properties.  
CC By analysing the 3-dimensional crystalline structure of the pertussis  
CC holotoxin, functional (aa) which affect biological properties of the  
CC pertussis holotoxin can be identified. This can be used to predict (aa)  
CC which contribute to the toxicity of the holotoxin to produce  
CC immunoprotective, genetically detoxified analogues of pertussis  
CC holotoxin. The present sequence represents an ADP-ribosylating toxin  
CC B-subunit peptide used in the exemplification of the present  
CC invention.

XX SQL Sequence 93 AA;

Query Match 100.0%; Score 62; DB 20; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.0003;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
Db 40 vevpgsqhdsq 51

RESULT 13

AAW95226  
ID AAW95226 standard; peptide; 93 AA.

XX

AAW95226;  
 16-MAR-1999 (first entry)  
 E. coli heat-labile toxin (LT) beta-subunit sequence.  
 Pertussis holotoxin; PT; modified; effector; toxicity; cell binding;  
 enzyme; ADP-ribosylation; toxin; crystalline; X-ray crystallography;  
 structural analysis; interacting site; mitogenicity; adjuvanticity;  
 heat-labile; LT.  
 Escherichia coli.  
 US5856122-A.  
 05-JAN-1999.  
 22-AUG-1994; 94US-0292968.  
 22-AUG-1994; 94US-0292968.  
 24-AUG-1993; 93US-0110947.  
 31-MAY-1994; 94US-0251121.  
 (UYAL-) UNIV ALBERTA.  
 Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;  
 Oomen RP, Read RJ, Stein PE;  
 WPI; 1999-105104/09.  
 Modifications to e.g. enzymatic activity, mitogenicity and cell  
 binding of pertussis holotoxin - by identifying interaction sites of  
 a molecule with crystalline toxin and modifying the identified site  
 Example 3; Fig 5; 40pp; English.  
 The invention relates to methods of preparing a pertussis holotoxin (PT)  
 having a modified biological activity. One method comprises identifying  
 at least 1 site in a PT that interacts with a molecule that is capable of  
 forming a complex with the holotoxin and which molecule is an effector  
 molecule which is an adenine nucleotide and which site contributes to  
 toxicity, cell binding or enzymatic activity of PT. The functional  
 interacting site(s) are identified by analysing the three dimensional  
 structure of crystalline PT, determined by X-ray crystallography. The  
 identified interacting site(s) are modified to alter toxicity, cell  
 binding or enzyme activity of the PT. The methods can be used to alter a  
 biological activity such as toxicity, enzymatic activity, mitogenicity,  
 cell binding and adjuvanticity of the PT. The three-dimensional structure  
 of PT have functional and/or structural resemblance to other bacterial  
 toxins such as diphtheria toxin (DT), Pseudomonas exotoxin A (ETA), the  
 heat-labile toxin of E. coli (LT) and verotoxin-1 (VT). The present  
 sequence represents the beta-subunit of LT toxin.  
 Sequence 93 AA;  
 Query Match 100.0%; Score 62; DB 20; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 0.0003;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VEVPGSQHDSQ 12  
 Db 40 vevpgsqhdsq 51  
 RESULT 14  
 AAY68365  
 ID AAY68365 standard; Peptide; 93 AA.  
 AC AAY68365;  
 XX  
 DT 17-APR-2000 (first entry)  
 XX

DE Heat labile toxin B subunit SEQ ID NO:26.  
 XX Pertussis toxin; heat labile toxin; verotoxin; exotoxin A; holotoxin;  
 KW diphtheria toxin; ADP-ribosylating toxin; mannose binding protein;  
 KW infection; crystal structure; X-ray crystallography; detoxification;  
 KW immunogenic.  
 XX Escherichia coli.  
 OS US6018022-A.  
 PN 25-JAN-2000.  
 XX  
 PD  
 XX  
 PF 06-JUN-1995; 95US-0467976.  
 XX  
 PR 22-AUG-1994; 94US-0292968.  
 PR 24-AUG-1993; 93US-0110947.  
 PR 31-MAY-1994; 94US-0251121.  
 XX (CONN-) CONNAUGHT LAB LTD.  
 PA (UYAL-) UNIV ALBERTA.  
 XX  
 PI Read RJ, Stein PE, Cockle SA, Loosmore S, Klein MH, Armstrong GD;  
 PI Hazes B, Oomen RP;  
 XX  
 DR WPI; 2000-136703/12.  
 XX  
 PT Crystalline form of isolated pertussis holotoxin useful in studying  
 PT proteins which have functional resemblance -  
 XX  
 PS Example 3; Fig 5; 42pp; English.  
 XX  
 CC The present invention describes a crystalline form of isolated  
 CC pertussis holotoxin, in which the molecules of pertussis toxin have  
 CC a three dimensional structure represented in the specification,  
 CC complexed with a polysaccharide molecule capable of forming a complex  
 CC with the holotoxin. The crystalline form of the pertussis holotoxin  
 CC can be used in a comparison with other proteins which have functional  
 CC resemblance to pertussis holotoxin with the aim of modifying other  
 CC proteins. Identifying the unknown sites of toxicity by comparison  
 CC with the three dimensional structure of pertussis holotoxin provides a  
 CC technique for detoxification of toxins to produce useful immunogenic  
 CC but non-toxic analogues. It can also be used as a primary standard to  
 CC measure the quantity, purity or efficacy of less pure compositions  
 CC containing pertussis toxin. AAY68340 to AAY68385 represent peptides  
 CC used in the exemplification of the present invention.  
 XX Sequence 93 AA;  
 SQ  
 Query Match 100.0%; Score 62; DB 21; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 0.0003;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VEVPGSQHDSQ 12  
 Db 40 vevpgsqhdsq 51  
 RESULT 15  
 AAB66239  
 ID AAB66239 standard; Protein; 93 AA.  
 XX  
 AC AAB66239;  
 XX  
 DT 03-APR-2001 (first entry)  
 XX  
 DE E coli verotoxin-1 B subunit SEQ ID NO: 26.  
 XX  
 KW Pertussis toxin; crystal structure; whooping cough; biological activity;  
 KW lymphocytosis-promoting factor; histamine-sensitising factor;  
 KW islet-activating protein.  
 XX

OS Escherichia coli.  
 XX  
 PN US6168928-B1.  
 XX  
 PD 02-JAN-2001.  
 XX  
 PF 21-MAY-1998; 98US-0082514.  
 XX  
 PR 22-AUG-1994; 94US-0292968.  
 PR 24-AUG-1993; 93US-0110947.  
 PR 31-MAY-1994; 94US-0251121.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 XX  
 PI Read RJ, Cockle SA, Oomen RP, Loosmore S, Klein MH, Armstrong GD;  
 PI Haze B, Stein PE;  
 XX  
 DR WPI; 2001-122260/13.  
 XX  
 PT Modifying pertussis holotoxin to produce detoxified PT analogs,  
 PT comprising analyzing crystalline structure of toxin, to identify sites  
 PT of toxicity, cell binding or enzyme activity of PT and modifying  
 PT identified site  
 XX  
 PS Example 3; Fig 5; 41pp; English.  
 XX  
 CC The present invention provides a method for producing a pertussis toxin  
 CC (also designated lymphocytosis-promoting factor, histamine-sensitising  
 CC factor and islet activating protein) with a modified biological activity,  
 CC involving analysing the crystal structure of the protein to identify  
 CC active sites which can then be modified. This may lead to an alteration  
 CC in the toxicity, cell binding or enzyme activity of the toxin. This can  
 CC be used in the production of immunoprotective analogues of pertussis  
 CC toxin. Pertussis toxin is the cause of whooping cough following infection  
 CC by Bordetella pertussis.  
 XX  
 SQ Sequence 93 AA;

Query Match 100.0%; Score 62; DB 22; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 0.0003;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
 Db 40 vevpgsqhidsq 51  
 |||||

RESULT 16  
 AAP50340  
 ID AAP50340 standard; protein; 103 AA.  
 XX  
 AC AAP50340;  
 XX  
 DT 01-DEC-1991 (first entry)  
 XX  
 DE Sequence of sub-unit B of cholera toxin.  
 XX  
 KW Vaccine; cholera; heat-labile E.coli toxin.  
 XX  
 OS Vibrio cholera.  
 XX  
 PH Key Location/Qualifiers  
 FT Region 50..64  
 FT /note= "claimed"  
 FT Region 8..20  
 FT /note= "claimed"  
 FT Region 45..64  
 FT /note= "claimed"  
 XX  
 PN DE3430894-A.  
 XX  
 PD 14-MAR-1985.

XX  
 PF 22-AUG-1984; 84DE-3430894.  
 XX  
 PR 23-AUG-1983; 83IL-0069558.  
 XX  
 PA (YEDA ) YEDA RES & DEV CO LTD.  
 XX  
 PI Sela M, Arnon R, Jacob CO;  
 XX  
 DR WPI; 1985-069683/12.  
 XX  
 PT Vaccines against cholera and heat-labile E. coli toxin - contg.  
 PT cholera toxin fragment coupled to carrier  
 XX  
 PS Example; Fig 1; 24pp; German.  
 XX  
 CC The inventors claim vaccines against cholera and heat-labile E.coli  
 CC toxin contg. cholera toxin fragment coupled to carrier. The toxin is  
 CC esp. the fragments defined in FT, above.  
 XX  
 SQ Sequence 103 AA;

Query Match 100.0%; Score 62; DB 6; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 0.00034;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
 Db 50 vevpgsqhidsq 61  
 |||||

RESULT 17  
 AAW04857  
 ID AAW04857 standard; Protein; 103 AA.  
 XX  
 AC AAW04857;  
 XX  
 DT 21-FEB-1997 (first entry)  
 XX  
 DE Synthetic cholera toxin B subunit.  
 XX  
 KW Bordetella pertussis; whooping cough; recombinant construct;  
 KW cholera toxin B subunit; enzyme; antigen; immunogen; allergen;  
 KW enzyme inhibitor; hormone; lymphokine; immunoglobulin; toxin;  
 KW structural protein; receptor; heterologous gene; leader; promoter.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9626282-A1.  
 XX  
 PD 29-AUG-1996.  
 XX  
 PF 23-FEB-1996; 96WO-CA00107.  
 XX  
 PR 23-FEB-1995; 95US-0393334.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 XX  
 PI Klein MH, Loosmore SM, Yacoob RK, Zealey GR;  
 XX  
 DR WPI; 1996-425088/42.  
 DR N-PSDB; AAT38038.  
 XX  
 PT Recombinant constructs for expressing and opt. secreting proteins in  
 PT Bordetella - comprise Bordetella promoter coupled to non-Bordetella,  
 PT esp. cholera B toxin, gene or coupled to non-Bordetella leader and  
 PT gene of interest  
 XX  
 PS Example 1; Figure 1; 61pp; English.  
 XX  
 CC Recombinant constructs comprising a promoter functional in  
 CC Bordetella operatively linked to a heterologous gene or a non-

CC Bordetella leader sequence for secretion of a gene product which may  
 CC or may not be of Bordetella origin, can be used for the expression  
 CC in Bordetella of enzymes, antigens, immunogens, allergens, enzyme  
 CC inhibitors, hormones, lymphokines, immunoglobulins or their  
 CC fragments, toxins, mammalian proteins, structural proteins or  
 CC receptors. The Bordetella strains are particularly engineered to  
 CC express the cholera toxin B subunit (this sequence). The promoters  
 CC used in the constructs are selected from the Bordetella pertussis  
 CC tox, fha promoters or the high molecular weight (hmw) outer membrane  
 CC promoter of non typable Haemophilus influenzae; leaders used in the  
 CC constructs are selected from the cholera toxin B leader (CTB-L), the  
 CC pertussis toxin subunit S1 leader (S1-L) and the pertussin pertactin  
 CC leader (PRN-L); and genes used in the constructs are selected from a  
 CC novel synthetic cholera toxin B gene (ctb) and the hmw1 and hmw2  
 CC genes of Haemophilus influenzae.  
 XX  
 XX Sequence 103 AA;

Query Match 100.0%; Score 62; DB 17; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 0.00034;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSOHDSQ 12  
 |||||

Db 50 vevpgsqhdsq 61

# RESULT 18

AAW06606  
 ID AAR94939 standard; Protein; 103 AA.

AC AAR94939;

DT 31-OCT-1996 (first entry)

DE Heat labile enterotoxin B subunit (LT-B) E.coli.

KW Toxin; subunit; vaccine; transgenic plant; immunogen; antigen;  
 KW adjuvant; immunisation.

OS Escherichia coli.

PN WO9612801-AL.

PD 02-MAY-1996.

PF 24-OCT-1995; 95WO-US13376.

PR 24-OCT-1994; 94US-0328716.

XX (TULA ) TULANE EDUCATIONAL FUND.  
 XX (TEXA ) UNIV TEXAS A & M SYSTEM.

PI Arntzen CJ, Clements JD, Haq TA, Mason HS;

XX WPI; 1996-230602/23.

DR N-PSDB; AAT18799, AAT18800.

XX Transgenic plants contg. E. coli heat labile enterotoxin subunits  
 XX used as oral vaccines for animals which consume the plant

PS Disclosure; Page 100-101; 130pp; English.

XX A transgenic plant comprising or expressing a DNA sequence encoding  
 XX an immunogen agent can be used as an oral vaccine for animals.  
 XX The vaccine is administered by the oral consumption of the plant and  
 XX provides the first known functional method for immunising animals  
 XX using transgenic plants, where the plants express bacterial antigens  
 XX that act as both immunogens and adjuvants. The method provides an  
 XX inexpensive production and delivery system for such antigens to  
 XX animals. This is the LT-B Escherichia coli toxin subunit and its  
 XX coding sequence was used in the construction of such a transgenic

CC plant. The immunogenic agent preferably comprises the LT-B or CT-B  
 CC (cholera toxin B subunit) or optionally LT-A or CT-A.  
 XX  
 XX Sequence 103 AA;

Query Match 100.0%; Score 62; DB 17; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 0.00034;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSOHDSQ 12

Db 50 vevpgsqhdsq 61

# RESULT 19

AAW06606

ID AAW06606 standard; Protein; 103 AA.

AC AAW06606;

DT 06-AUG-1997 (first entry)

DE Cholera toxin B subunit/heat labile enterotoxin B subunit hybrid.

XX Heat labile; tolerant; stable; hybrid; cholera toxin B subunit; CTB;  
 KW heat labile enterotoxin B subunit; LTB; vaccine; immunisation;  
 KW immunogen; antigen; broad spectrum; cross reactive; diarrhoea;  
 KW vomiting; food poisoning.

XX Chimeric - Vibrio cholerae.

OS Chimeric - Enterotoxigenic Escherichia Coli.

XX

FT Key Location/Qualifiers

FT Misc-difference 1

FT /label= substitution

FT /note= "wild-type Thr replaced by Ala"

FT Misc-difference 94

FT /label= substitution

FT /note= "wild-type His replaced by Asn"

FT Misc-difference 95

FT /label= substitution

FT /note= "wild-type Ala replaced by Ser"

XX WO9634893-AL.

XX 07-NOV-1996.

XX 02-MAY-1996; 96WO-SE00570.

XX 05-MAY-1995; 95SE-0001682.

XX (HOLM/) HOLMGREN J.

XX (LEBE/) LEBENS M R.

XX Holmgren J, Lebens MR;

XX WPI; 1996-506108/50.

XX N-PSDB; AAT43576.

XX Heat-labile enterotoxin B sub-unit and cholera toxin B sub-unit  
 XX hybrid protein - opt. fused to immunogenic sequence for use in  
 XX vaccines against enterotoxin-induced illness

PS Claim 3; Page -; 32pp; English.

XX AAW06606 is a mature cholera toxin B subunit (CTB)/heat labile  
 CC enterotoxin B subunit (LTB) hybrid molecule. In mature CTB protein,  
 CC certain amino acids (aa) were replaced with corresponding aa from  
 CC heat-labile enterotoxin B subunit (LTB). The specific amino acid  
 CC substitutions impart LTB-specific epitope characteristics to  
 CC immunogenic mature CTB. The hybrid molecules have increased  
 CC cross-reactivity and are suitable for a broad spectrum vaccine to

CC protect against enterotoxigenic illness. Immunogenic proteins  
 CC comprising the hybrid molecules can be used to treat, or in a  
 CC vaccine to prevent, enterotoxin-induced illness, e.g. diarrhoea  
 CC and vomiting, in humans and animals.  
 CC Note - This sequence does not appear in the specification, it is  
 CC a claimed mutant sequence of mature cholera toxin B subunit (see  
 CC AAW06605).

XX Sequence 103 AA;

Query Match 100.0%; Score 62; DB 17; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 0.00034;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
 |||||  
 Db 50 vevpgsqhdsq 61

#### RESULT 20

AAW06607  
 ID AAW06607 standard; Protein; 103 AA.

XX AC AAW06607;

DT 06-AUG-1997 (first entry)

XX DE Cholera toxin B subunit/heat labile enterotoxin B subunit hybrid.

XX KW Heat labile; tolerant; stable; hybrid; cholera toxin B subunit; CTB;  
 KW heat labile enterotoxin B subunit; LTB; vaccine; immunisation;  
 KW immunogen; antigen; broad spectrum; cross reactive; diarrhoea;  
 KW vomiting; food poisoning.

XX OS Chimeric - Vibrio cholerae.

OS OS Chimeric - Enterotoxigenic Escherichia Coli.

XX FH Key Location/Qualifiers

FT Misc-difference 1..25

FT /label- substitution

FT /note- "the first 25 amino acids of mature  
 wild-type cholera toxin B subunit are  
 replaced with the first 25 amino acids  
 of mature enterotoxin B subunit"

XX PN WO9634893-A1.

XX XX 07-NOV-1996.

XX PF 02-MAY-1996; 96WO-SE00570.

XX PR 05-MAY-1995; 95SE-0001682.

XX PA (HOLM/) HOLMGREN J.

PA PA (LEBE/) LEBENS M R.

XX PI Holmgren J, Lebens MR;

XX DR WPI; 1996-506108/50.

DR N-PSDB; AAT43577.

XX Heat-labile enterotoxin B sub-unit and cholera toxin B sub-unit  
 PT hybrid protein - opt. fused to immunogenic sequence for use in  
 PT vaccines against enterotoxin-induced illness

XX PS Claim 4; Page -; 32pp; English.

XX AAW06607 is a mature cholera toxin B subunit (CTB)/heat labile  
 CC enterotoxin B subunit (LTB) hybrid molecule. In mature CTB protein,  
 CC certain amino acids (aa) were replaced with corresponding aa from  
 CC heat-labile enterotoxin B subunit (LTB). The specific amino acid  
 CC substitutions impart LTB-specific epitope characteristics to

CC immunogenic mature CTB. The hybrid molecules have increased  
 CC cross-reactivity and are suitable for a broad spectrum vaccine to  
 CC protect against enterotoxigenic illness. Immunogenic proteins  
 CC comprising the hybrid molecules can be used to treat, or in a  
 CC vaccine to prevent, enterotoxin-induced illness, e.g. diarrhoea  
 CC and vomiting, in humans and animals.

CC Note - This sequence does not appear in the specification, it is  
 CC a claimed mutant sequence of mature cholera toxin B subunit (see  
 CC AAW06605).

XX SQ Sequence 103 AA;

Query Match 100.0%; Score 62; DB 17; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 0.00034;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
 |||||  
 Db 50 vevpgsqhdsq 61

#### RESULT 21

AAW80808

ID AAW80808 standard; protein; 103 AA.

XX AC AAW80808;

DT 29-JAN-1999 (first entry)

XX DE Amino acid sequence of the wild type cholera toxin B subunit.

XX KW Cholera toxin B subunit; nontoxic subunit; adjuvant; coadministration;  
 KW antigen; bird; animal; mucosal; vaccine.

XX OS Vibrio cholerae.

XX PN WO9845324-A1.

XX PD 15-OCT-1998.

XX PF 03-APR-1998; 98WO-US06725.

XX PR 04-APR-1997; 97US-0043410.

XX PA (KIYO/) KIYONO H.

PA PA (MCGH/) MCGHEE J R.

PA PA (TAKE/) TAKEDA Y.

PA PA (UABR-) UAB RES FOUND.

PA PA (YAMA/) YAMAMOTO S.

XX PI Kiyono H, Mcghee JR, Takeda Y, Yamamoto S;

XX DR WPI; 1998-594478/50.

XX New mutant cholera toxin selected from a group comprising nontoxic  
 PT subunits/derivatives - effective as an adjuvant when coadministered  
 PT with an antigen to birds and mammals

XX PS Disclosure; Fig 1B; 43pp; English.

XX This is the amino acid sequence of the cholera toxin B subunit used in  
 CC the method of the invention involving the use of nontoxic subunits as  
 CC an effective adjuvant in coadministration of an antigen to birds  
 CC and animals. In addition to the use of the toxin as an mucosal  
 CC adjuvant, it also provides a vaccine comprising the toxin, an  
 CC immunogenic amount of an antigen, and a pharmaceutically acceptable  
 CC carrier. The toxin can be used with single/multiple vaccines, and it  
 CC enables the possibility for commercial mucosal adjuvants for use in  
 CC humans, since these are more effective and safer than vaccines  
 CC administered subcutaneously.

XX SQ Sequence 103 AA;

```

Query Match      100.0%; Score 62; DB 19; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.00034;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHIDSQ 12
Db 50 vevpgsqhidsq 61

RESULT 22
AAR04163
ID AAR04163 standard; protein; 118 AA.
XX
AC AAR04163;
XX
DT 10-SEP-1990 (first entry)
DE Cholera Toxin B-subunit.
XX
KW cholera toxin B-subunit; hybrid protein; heterologous IgA active antigen.
XX
OS synthetic.
XX
FH Key Location/Qualifiers
FT misc_difference 18..18 /*label= His or Tyr
FT region 1..11 /*label= signal peptide
FT /*note= absent from mature protein
XX
PN W09003437-A.
XX
PD 05-APR-1990.
XX
PF 27-SEP-1989; 89WO-0000495.
XX
PR 27-SEP-1988; 88FR-0012627.
XX
PA (UYLI-) L'UNIVERSITE DE L'ETAT A LIEGE.
XX
PI L'Hoir C, Renard A, Martial J;
XX
DR WPI: 1990-132273/17.
DR N-NSDB; Q04046.
XX
PT New hybrid protein, useful in vaccines -
PT contains cholera toxin b subunit and heterologous IgA active
PT antigenic sequence.
XX
PS Disclosure; ; pp; French.
XX
CC Mature cholera toxin B-subunit is obtained when the signal peptide is
CC cleaved off. There is an Ochre codon at position 343-5; the sequence
CC downstream from it is part of a plasmid.
XX
SQ Sequence 118 AA;

Query Match      100.0%; Score 62; DB 11; Length 118;
Best Local Similarity 100.0%; Pred. No. 0.00039;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHIDSQ 12
Db 61 vevpgsqhidsq 72

RESULT 23
AAP93561
ID AAP93561 standard; protein; 124 AA.
XX

```

```

AC AAP93561;
XX
DT 06-JUN-1990 (first entry)
XX
DE B subunit of the heat-labile enterotoxin (LT-B) derived from E. coli.
XX
KW B subunit; heat-labile enterotoxin; LT-B; Escherichia coli; malaria;
KW circumsporozoite protein; fusion protein; live recombinant vaccine;
KW Salmonella; epitope.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT Peptide 1..22 /note="Signal peptide"
FT Protein 23..124 /note="Mature LT-B"
XX
PN W08902924-A.
XX
PD 06-APR-1989.
XX
PF 30-SEP-1988; 88WO-US03376.
XX
PR 02-OCT-1987; 87US-0104735.
XX
PA (PRAX-) PRAXIS BIOLOGICS IN.
XX
PI Brey RN, Majarian WR, Pillai S, Hockmeyer WT;
XX
DR WPI: 1989-114399/15.
DR N-PSDB; AAN90747.
XX
PT Live recombinant vaccine for malaria -
PT comprising attenuated entero-invasive bacterium contg. DNA
PT encoding epitope of malaria parasite
XX
PS Fig 3; p. 3/17; 105pp; English.
XX
CC In the patent, the DNA encoding LT-B is expressed as part of a fusion
CC protein with an epitope of a malaria parasite, eg Region I or Region II
CC or a repeat region of circumsporozoite protein antigen (CS) (AAP93560)
CC from Plasmodium berghei. Pref. the fusion gene is inserted into
CC attenuated Salmonella enteritidis under the left promoter control of
CC lambda. Such bacteria can multiply in the host without causing disease or
CC disorder and express CS that will induce a protective immune response
CC against malaria and can be used in vaccines. Such vaccines can be
CC multivalent.
XX
SQ Sequence 124 AA;

Query Match      100.0%; Score 62; DB 10; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHIDSQ 12
Db 71 vevpgsqhidsq 82

RESULT 24
AAW06605
ID AAW06605 standard; Protein; 124 AA.
XX
AC AAW06605;
XX
DT 06-AUG-1997 (first entry)
DE Cholera toxin B subunit, used for hybrid immunogenic toxin production.
XX
KW Heat labile; tolerant; stable; hybrid; cholera toxin B subunit; CTB;
KW heat labile enterotoxin B subunit; LTB; vaccine; immunisation;

```

KW immunogen; antigen; broad spectrum; cross reactive; diarrhoea;  
 KW vomiting; food poisoning.

XX Vibrio cholerae.

XX Key Location/Qualifiers  
 FH Peptide 1..21  
 FT /label= sig\_peptide  
 FT Protein 22..124  
 FT /label= mat\_protein

XX WO9634893-A1.

PN 07-NOV-1996.

XX 02-MAY-1996; 96WO-SE00570.

XX 05-MAY-1995; 95SE-0001682.

XX (HOLM/) HOLMGREN J.

XX (LEBE/) LEBENS M R.

XX Holmgren J, Lebens MR;

XX WPI; 1996-506108/50.

XX N-PSDB; AAT43575.

XX Heat-labile enterotoxin B sub-unit and cholera toxin B sub-unit  
 PT hybrid protein - opt. fused to immunogenic sequence for use in  
 PT vaccines against enterotoxin-induced illness

XX Disclosure; Fig 1; 32pp; English.

XX AAM06605 is the full (including the signal peptide) length cholera  
 CC toxin B subunit (CTB), this sequence is described as unpublished in  
 CC the specification. The mature CTB protein was used to create hybrid  
 CC mutants, in which certain amino acids (aa) of CTB were replaced with  
 CC corresponding aa from heat-labile enterotoxin B subunit (LTB), see  
 CC AAM06606 and AAM06607. The specific amino acid substitutions impart  
 CC LTB-specific epitope characteristics to immunogenic mature CTB. The  
 CC hybrid molecules have increased cross-reactivity and are suitable  
 CC for a broad spectrum vaccine to protect against enterotoxigenic  
 CC illness. Immunogenic proteins comprising the hybrid molecules can be  
 CC used to treat, or in a vaccine to prevent, enterotoxin-induced illness,  
 CC e.g. diarrhoea and vomiting, in humans and animals.

XX Sequence 124 AA;

Query Match 100.0%; Score 62; DB 17; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.00041;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VEVPGSQHDSQ 12  
 Db 71 vevpgsqhidsq 82  
 |||||

RESULT 25

AAW59770

ID AAW59770 standard; Protein; 124 AA.

XX AAW59770;

XX 12-OCT-1998 (first entry)

XX Amino acid sequence of E. coli LTB.

XX Beta-subunit of heat labile enterotoxin; LTB; fusion protein; vaccine;  
 KW immunogen; antigen; inhibitor; fertility; follicle stimulating hormone;  
 KW FSH; sperm; ova; immune response.

XX Escherichia coli.

XX WO9821344-A1.

XX 22-MAY-1998.

XX 12-NOV-1997; 97WO-US20584.

XX 12-NOV-1996; 96US-0747410.

XX (UNMS ) UNIV MICHIGAN STATE.

XX Bagdasarian M, Ireland J;

XX WPI; 1998-297947/26.

XX N-PSDB; AAV41573.

XX New nucleic acid encoding fusion of antigenic peptide and  
 PT enterotoxin sub-unit - useful as vaccinating immunogen, particularly  
 PT for increasing animal fertility by inducing antibodies against  
 PT inhibin

XX Disclosure; Fig 9; 56pp; English.

XX This is the amino acid sequence of Escherichia coli beta-subunit of  
 CC heat labile enterotoxin (LTB). It is used in the method of the  
 CC invention to create fusion proteins which are useful as vaccinating  
 CC immunogens. The fusion proteins are useful in vaccines, specifically  
 CC where the antigenic peptide is an inhibin fragment for increasing the  
 CC fertility of an animal (by increasing levels of follicle stimulating  
 CC hormone (FSH) or production of sperm or ova), but more generally for  
 CC inducing an immune response against the antigenic peptide. Vaccines  
 CC are particularly administered orally, e.g. fusion protein is expressed  
 CC in edible plants or animals.

XX Sequence 124 AA;

Query Match 100.0%; Score 62; DB 19; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.00041;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VEVPGSQHDSQ 12  
 Db 71 vevpgsqhidsq 82  
 |||||

Search completed: July 16, 2001, 16:35:41  
 Job time: 205 sec





*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 16:36:18 ; Search time 30.3 Seconds  
(without alignments)  
7.978 Million cell updates/sec

Title: US-09-786-648-3  
Perfect score: 62  
Sequence: 1 VEPGSHIDSQ 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 2014635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/2/iaa/PTUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	100.0	93	2	US-08-292-968-26
2	62	100.0	93	2	US-08-467-974-26
3	62	100.0	93	2	US-08-467-536-26
4	62	100.0	93	3	US-08-467-976-26
5	62	100.0	93	4	US-09-082-514-26
6	62	100.0	102	3	US-08-952-337-5
7	62	100.0	102	3	US-08-952-337-6
8	62	100.0	103	2	US-08-472-171-2
9	62	100.0	103	2	US-08-894-526-2
10	62	100.0	103	2	US-09-013-047-2
11	62	100.0	103	4	US-09-374-597-2
12	62	100.0	103	4	US-09-191-852-21
13	62	100.0	103	5	PT-US95-13376-21
14	62	100.0	123	3	US-08-952-337-1
15	62	100.0	123	3	US-08-952-337-2
16	62	100.0	124	2	US-08-747-410-2
17	62	100.0	371	2	US-08-829-028A-6
18	56	90.3	124	1	US-08-449-045C-4
19	56	90.3	124	2	US-08-435-605A-12
20	56	90.3	124	6	5223610-3
21	39	62.9	346	2	US-08-602-359A-34
22	37	59.7	448	2	US-08-878-989-2
23	37	59.7	448	4	US-09-272-796-2
24	36	58.1	855	2	US-09-027-337-2
25	34	54.8	459	6	5194375-6
26	34	54.8	775	2	US-08-714-070A-1
27	33	53.2	321	3	US-09-039-609-4

28	33	53.2	458	3	US-09-039-609-2	Sequence 2, Appli
29	33	53.2	845	1	US-08-416-950-11	Sequence 11, Appl
30	33	53.2	845	2	US-08-469-830-11	Sequence 11, Appl
31	32	51.6	282	1	US-07-672-304-3	Sequence 3, Appli
32	32	51.6	297	1	US-08-011-398B-4	Sequence 4, Appli
33	32	51.6	297	1	US-08-464-051-4	Sequence 4, Appli
34	32	51.6	297	2	US-08-462-498-4	Sequence 4, Appli
35	32	51.6	297	3	US-08-554-385-3	Sequence 3, Appli
36	32	51.6	360	1	US-08-674-612-3	Sequence 3, Appli
37	32	51.6	360	1	US-08-469-421-14	Sequence 14, Appl
38	32	51.6	360	1	US-08-250-975-14	Sequence 14, Appl
39	32	51.6	360	2	US-08-920-296-3	Sequence 3, Appli
40	32	51.6	360	2	US-08-605-002A-14	Sequence 14, Appl
41	32	51.6	360	2	US-08-950-449A-14	Sequence 14, Appl
42	32	51.6	360	3	US-08-554-385-20	Sequence 20, Appl
43	32	51.6	360	4	US-09-286-904-2	Sequence 2, Appli
44	32	51.6	360	4	US-09-286-904-46	Sequence 46, Appl
45	32	51.6	360	4	US-09-025-580-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-292-968-26  
; Sequence 26, Application US/08292968  
; Patent No. 5856122  
; GENERAL INFORMATION:  
; APPLICANT: READ, Randy J.  
; APPLICANT: STEIN, Penelope E.  
; APPLICANT: COCKLE, Stephen A.  
; APPLICANT: COHEN, Raymond P.  
; APPLICANT: LOOMORE, Sheena  
; APPLICANT: KLEIN, Michel H.  
; APPLICANT: ARMSTRONG, Glen D.  
; APPLICANT: HAZES, Bart  
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/292,968  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/110,947  
; FILING DATE: 24-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/251,121  
; FILING DATE: 31-MAY-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-388  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 93 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-08-292-968-26

Query Match 100.0%; Score 62; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.00027;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
| | | | | | | | | |  
Db 40 VEVPGSQHDSQ 51

RESULT 2

US-08-467-974-26  
; Sequence 26, Application US/08467974  
; Patent No. 5965385  
; GENERAL INFORMATION:  
; APPLICANT: READ, Randy J.  
; APPLICANT: STEIN, Penelope E.  
; APPLICANT: COCKLE, Stephen A.  
; APPLICANT: OOMEN, Raymond P.  
; APPLICANT: LOOSMORE, Sheena  
; APPLICANT: KLEIN, Michel H.  
; APPLICANT: ARMSTRONG, Glen D.  
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467,974  
; FILING DATE:  
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/467,536  
; FILING DATE: 06-JUN-1995  
; APPLICATION NUMBER: US 08/292,968  
; FILING DATE: 22-AUG-1994  
; APPLICATION NUMBER: US 08/251,121  
; FILING DATE: 31-MAY-1994

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/110,947  
; FILING DATE: 24-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-454 MIS:vg

TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 93 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-08-467-974-26

Query Match 100.0%; Score 62; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.00027;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
| | | | | | | | | |  
Db 40 VEVPGSQHDSQ 51

RESULT 3

US-08-467-536-26  
; Sequence 26, Application US/08467536  
; Patent No. 5977304  
; GENERAL INFORMATION:  
; APPLICANT: READ, Randy J.  
; APPLICANT: STEIN, Penelope E.  
; APPLICANT: COCKLE, Stephen A.  
; APPLICANT: OOMEN, Raymond P.  
; APPLICANT: LOOSMORE, Sheena  
; APPLICANT: KLEIN, Michel H.  
; APPLICANT: ARMSTRONG, Glen D.  
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467,536  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/292,968  
; FILING DATE: 22-AUG-1994  
; APPLICATION NUMBER: US 08/251,121  
; FILING DATE: 31-MAY-1994  
; APPLICATION DATA:  
; APPLICATION NUMBER: US 08/110,947  
; FILING DATE: 24-AUG-1993

ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-455 MIS:vg  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 93 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-08-467-536-26

Query Match 100.0%; Score 62; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.00027;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
| | | | | | | | | |  
Db 40 VEVPGSQHDSQ 51

RESULT 4

US-08-467-976-26

```

; Sequence 26, Application US/08467976
; Patent No. 6018022
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J.
; APPLICANT: STEIN, Penelope E.
; APPLICANT: COCKLE, Stephen A.
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: KLEIN, Michel H.
; APPLICANT: ARMSTRONG, Glen D.
; APPLICANT: HAZES, Bart
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,976
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,968
; FILING DATE: 22-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/251,121
; FILING DATE: 31-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/110,947
; FILING DATE: 24-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-453 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-467-976-26

```

```

Query Match 100.0%; Score 62; DB 3; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 VEVPGSQHDSQ 12
   |||||
Db 40 VEVPGSQHDSQ 51

```

```

RESULT 5
US-09-082-514-26
; Sequence 26, Application US/09082514
; Patent No. 6168928
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J.
; APPLICANT: STEIN, Penelope E.
; APPLICANT: COCKLE, Stephen A.
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.

```

```

; APPLICANT: ARMSTRONG, Glen D.
; APPLICANT: HAZES, Bart
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/082,514
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,968
; FILING DATE: 24-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-810
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-082-514-26

```

```

Query Match 100.0%; Score 62; DB 4; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 VEVPGSQHDSQ 12
   |||||
Db 40 VEVPGSQHDSQ 51

```

```

RESULT 6
US-08-952-337-5
; Sequence 5, Application US/08952337
; Patent No. 6019973
; GENERAL INFORMATION:
; APPLICANT: Holmgren, Jan
; APPLICANT: Lebens, Michael R.
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
; FILE REFERENCE: 3846/0D758
; CURRENT APPLICATION NUMBER: US/08/952,337
; CURRENT FILING DATE: 1998-01-05
; EARLIER APPLICATION NUMBER: PCT/SE96/00570
; EARLIER FILING DATE: 1996-05-02
; EARLIER APPLICATION NUMBER: SE 9501682-0
; EARLIER FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Vibrio cholerae
; US-08-952-337-5

```

Query Match 100.0%; Score 62; DB 3; Length 102;  
Best Local Similarity 100.0%; Pred. No. 0.0003;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
| | | | | | | | | | | | | |  
DB 49 VEVPGSQHDSQ 60

## RESULT 7

US-08-952-337-6  
; Sequence 6, Application US/08952337  
; Patent No. 6019973  
; GENERAL INFORMATION:  
; APPLICANT: Holmgren, Jan  
; APPLICANT: Lebens, Michael R.  
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE  
; FILE REFERENCE: 3846/00758  
; CURRENT APPLICATION NUMBER: US/08/952,337  
; CURRENT FILING DATE: 1998-01-05  
; EARLIER APPLICATION NUMBER: PCT/SE96/00570  
; EARLIER FILING DATE: 1996-05-02  
; EARLIER APPLICATION NUMBER: SE 9501682-0  
; EARLIER FILING DATE: 1995-05-05  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 102  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-08-952-337-6

Query Match 100.0%; Score 62; DB 3; Length 102;  
Best Local Similarity 100.0%; Pred. No. 0.0003;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
| | | | | | | | | | | | | |  
DB 49 VEVPGSQHDSQ 60

## RESULT 8

US-08-472-171-2  
; Sequence 2, Application US/08472171  
; Patent No. 5932714  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yacoub, Reza K.  
; APPLICANT: Zealey, Gavin R.  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Expression of Gene Products From  
; Genetically Manipulated Strains of Bordetella  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 330 University Avenue, Suite 701  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/472,171  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/393,334

; FILING DATE: 23-FEB-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-507 MIS:vg  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 416-595-1155  
; TELEFAX: 416-595-1163  
; TELEX: 065-24567 Simbas  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 103 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-472-171-2

Query Match 100.0%; Score 62; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.0003;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
| | | | | | | | | | | | | |  
DB 50 VEVPGSQHDSQ 61

RESULT 9  
US-08-894-526-2  
; Sequence 2, Application US/08894526  
; Patent No. 5942418  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M  
; APPLICANT: Yacoub, Reza K  
; APPLICANT: Zealey, Gavin R  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: EXPRESSION OF GENE PRODUCTS FROM  
; GENETICALLY MANIPULATED STRAINS OF BORDETELLA  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/894,526  
; FILING DATE: 01-DEC-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-724 MIS:jb  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 103 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-894-526-2

Query Match 100.0%; Score 62; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.0003;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
|||||  
Db 50 VEVPGSQHDSQ 61

## RESULT 10

US-09-013-047-2  
; Sequence 2, Application US/09013047  
; Patent No. 5998168  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yacoub, Reza K.  
; APPLICANT: Zealey, Gavin R.  
; TITLE OF INVENTION: Expression Of Gene Products From  
; TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 330 University Avenue, 6th Floor  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/013,047  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/472,171  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/393,334  
; FILING DATE: 23-FEB-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-507 MIS:vg  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 416-595-1155  
; TELEFAX: 416-595-1163  
; TELEX: 065-24567 Simbas  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 103 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-013-047-2

Query Match 100.0%; Score 62; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.0003;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
|||||  
Db 50 VEVPGSQHDSQ 61

## RESULT 11

US-09-374-597-2  
; Sequence 2, Application US/09374597  
; Patent No. 6140082  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.

; APPLICANT: Yacoub, Reza K.  
; APPLICANT: Zealey, Gavin R.  
; APPLICANT: Klein, Michael H.  
; TITLE OF INVENTION: Expression Of Gene Products From  
; TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 330 University Avenue, 6th Floor  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/374,597  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/393,334  
; FILING DATE: FEBRUARY 23, 1995  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-964  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 416-595-1155  
; TELEFAX: 416-595-1163  
; TELEX: 065-24567 Simbas  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 103 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-374-597-2

Query Match 100.0%; Score 62; DB 4; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.0003;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
|||||  
Db 50 VEVPGSQHDSQ 61

## RESULT 12

US-09-191-852-21  
; Sequence 21, Application US/09191852  
; Patent No. 6194560  
; GENERAL INFORMATION:  
; APPLICANT: Charles J. Arntzen, Hugh S. Mason, and Tariq A. Haq  
; TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fulbright & Jaworski L.L.P.  
; STREET: 1301 McKinney, Suite 5100  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77010  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/191.852  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13376  
FILING DATE: 24-OCT-1995  
APPLICATION NUMBER: 08/817,906  
FILING DATE: 04-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Fox, David L.  
REGISTRATION NUMBER: 40,612  
REFERENCE/DOCKET NUMBER: P01590US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-651-5151  
TELEFAX: 713-651-5246  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-09-191-852-21

Query Match 100.0%; Score 62; DB 4; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.0003;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
|||||

Db 50 VEVPGSQHDSQ 61

## RESULT 13

PCT-US95-13376-21  
Sequence 21, Application PC/TUS9513376  
GENERAL INFORMATION:  
APPLICANT: The Texas A&M University System  
APPLICANT: 310 Wisenbaker  
APPLICANT: College Station, Texas 77843-3369  
TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pravel, Hewitt, Kimball & Krieger  
STREET: 1177 West Loop South, 10th Floor  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77027-9095  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13376  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/328,716  
FILING DATE: 24-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jones, John W.  
REGISTRATION NUMBER: 31,380  
REFERENCE/DOCKET NUMBER: 36170/3P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-850-0909  
TELEFAX: 713-850-0165  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
PCT-US95-13376-21

Query Match 100.0%; Score 62; DB 5; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.0003;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
|||||

Db 50 VEVPGSQHDSQ 61

## RESULT 14

US-08-952-337-1  
Sequence 1, Application US/08952337  
Patent No. 6019973  
GENERAL INFORMATION:  
APPLICANT: Holmgren, Jan  
APPLICANT: Lebens, Michael R.  
TITLE OF INVENTION: HYBRID MOLECULES-BETWEEN HEAT-LABILE  
TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS  
FILE REFERENCE: 3846/0D758  
CURRENT APPLICATION NUMBER: US/08/952,337  
CURRENT FILING DATE: 1998-01-05  
EARLIER APPLICATION NUMBER: PCT/SE96/00570  
EARLIER FILING DATE: 1996-05-02  
EARLIER APPLICATION NUMBER: SE 9501682-0  
EARLIER FILING DATE: 1995-05-05  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 123  
TYPE: PRT  
ORGANISM: Vibrio cholerae  
US-08-952-337-1

Query Match 100.0%; Score 62; DB 3; Length 123;  
Best Local Similarity 100.0%; Pred. No. 0.00037;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
|||||

Db 70 VEVPGSQHDSQ 81

## RESULT 15

US-08-952-337-2  
Sequence 2, Application US/08952337  
Patent No. 6019973  
GENERAL INFORMATION:  
APPLICANT: Holmgren, Jan  
APPLICANT: Lebens, Michael R.  
TITLE OF INVENTION: HYBRID MOLECULES-BETWEEN HEAT-LABILE  
TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS  
FILE REFERENCE: 3846/0D758  
CURRENT APPLICATION NUMBER: US/08/952,337  
CURRENT FILING DATE: 1998-01-05  
EARLIER APPLICATION NUMBER: PCT/SE96/00570  
EARLIER FILING DATE: 1996-05-02  
EARLIER APPLICATION NUMBER: SE 9501682-0  
EARLIER FILING DATE: 1995-05-05  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 123  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-08-952-337-2

Query Match 100.0%; Score 62; DB 3; Length 123;  
Best Local Similarity 100.0%; Pred. No. 0.00037;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```

QY 1 VEVPGSHIDSQ 12
Db 70 VEVPGSHIDSQ 81

RESULT 16
US-08-747-410-2
; Sequence 2, Application US/08747410
; Patent No. 5993820
; GENERAL INFORMATION:
; APPLICANT: BAGDASARIAN, Michael
; APPLICANT: IRELAND, James
; TITLE OF INVENTION: CHIMERIC LT-B VACCINES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 3100 No. 5993820 West Center, 90 South Seventh St
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/747,410
; FILING DATE: 12-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kettelberger, Denise M
; REGISTRATION NUMBER: 33,924
; REFERENCE/DOCKET NUMBER: 11526.1-US-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612/371-5268
; TELEFAX: 612/332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 124 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
US-08-747-410-2

Query Match 100.0%; Score 62; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00037;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSHIDSQ 12
Db 71 VEVPGSHIDSQ 82

RESULT 17
US-08-829-026A-6
; Sequence 6, Application US/08829026A
; Patent No. 5837825
; GENERAL INFORMATION:
; APPLICANT: Meinersmann, Richard J.
; APPLICANT: Khoury, Christian A.
; TITLE OF INVENTION: Campylobacter Jejuni Flagellin-Escherichia Coli LT-B Fusion P

```

```

; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Janelle S. Graeter
; STREET: Room 411, Building 005, BARC-W
; CITY: Beltsville
; STATE: MD
; COUNTRY: USA
; ZIP: 20705
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/829,026A
; FILING DATE: 18-AUG-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Graeter, Janelle S.
; REGISTRATION NUMBER: 35,024
; REFERENCE/DOCKET NUMBER: 0106.97
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-504-5676
; TELEFAX: 301-504-5060
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 371 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-829-026A-6

Query Match 100.0%; Score 62; DB 2; Length 371;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSHIDSQ 12
Db 54 VEVPGSHIDSQ 65

RESULT 18
US-08-449-045C-4
; Sequence 4, Application US/08449045C
; Patent No. 5770203
; GENERAL INFORMATION:
; APPLICANT: Burnette, Neal W.
; APPLICANT: Kaslow, Harvey R.
; TITLE OF INVENTION: RECOMBINANT DNA-DERIVED CHOLERA TOXIN
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,045C
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/694,733
; FILING DATE: 02-MAY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/271,222

```

;; FILING DATE: 06-JUL-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Mazza, Richard J.  
;; REGISTRATION NUMBER: 27,657  
;; REFERENCE/DOCKET NUMBER: A-196C  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 124 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-449-045C-4

Query Match 90.3%; Score 56; DB 1; Length 124;  
Best Local Similarity 91.7%; Pred. No. 0.0044;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
| | | | | | | | | |  
DB 71 VEVPSQHDSQ 82

RESULT 19  
US-08-435-605A-12  
; Sequence 12, Application US/08435605A  
; Patent No. 5874287  
; GENERAL INFORMATION:  
; APPLICANT: Burnette, W. Neal  
; APPLICANT: Kaslow, Harvey R.  
; TITLE OF INVENTION: RECOMBINANT DNA-DERIVED CHOLERA TOXIN  
; TITLE OF INVENTION: SUBUNIT ANALOGS  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Amgen Inc.  
; STREET: 1840 De Havilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/435,605A  
; FILING DATE: 05-MAY-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mazza, Richard J.  
; REGISTRATION NUMBER: 27,657  
; REFERENCE/DOCKET NUMBER: A-196B  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 124 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-435-605A-12

Query Match 90.3%; Score 56; DB 2; Length 124;  
Best Local Similarity 91.7%; Pred. No. 0.0044;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
| | | | | | | | | |  
DB 71 VEVPSQHDSQ 82

RESULT 20

5223610-3  
; Patent No. 5223610  
; APPLICANT: Burton, Frank H.; Sutcliffe, Gregor  
; TITLE OF INVENTION: CHOLERA TOXIN GENE REGULATED BY GROWTH  
; HORMONE PROMOTER  
; NUMBER OF SEQUENCES: 18  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/528,852  
; FILING DATE: 18-MAY-1990  
; SEQ ID NO: 3;  
; LENGTH: 124  
5223610-3

Query Match 90.3%; Score 56; DB 6; Length 124;  
Best Local Similarity 91.7%; Pred. No. 0.0044;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
| | | | | | | | | |  
DB 71 VEVPSQHDSQ 82

RESULT 21  
US-08-602-359A-34  
; Sequence 34, Application US/08602359A  
; Patent No. 5942430  
; GENERAL INFORMATION:  
; APPLICANT: ROBERTSON, Daniel E.  
; APPLICANT: MURPHY, Dennis  
; APPLICANT: REID, John  
; APPLICANT: MAFFIA, Anthony  
; APPLICANT: LINK, Steven  
; APPLICANT: SWANSON, Ronald V.  
; APPLICANT: WARREN, Patrick V.  
; APPLICANT: KOSMOTKA, Anna  
; TITLE OF INVENTION: ESTERASES  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & RICHARDSON P. C.  
; STREET: 4225 EXECUTIVE SQUARE, STE 1400  
; CITY: LA JOLLA  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/602,359A  
; FILING DATE: February 16, 1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HAILE, LISA A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 09010/010001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-678-5070  
; TELEFAX: 619-678-5099  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 346 AMINO ACIDS  
; TYPE: AMINO ACID  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-08-602-359A-34

Query Match 62.9%; Score 39; DB 2; Length 346;  
Best Local Similarity 54.5%; Pred. No. 15;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 VEPVGSQHDS 11  
| | | | | : | : :  
Db 309 VEPVGSQHVEA 319

RESULT 22

US-08-878-989-2  
; Sequence 2, Application US/08878989  
; Patent No. 5885803  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Guegler, Karl G.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Goli, Surya K.  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN  
; TITLE OF INVENTION: KINASES  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/878,989  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0321 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 448 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: TBLYNOT01  
CLONE: 40194  
US-08-878-989-2

Query Match 59.7%; Score 37; DB 2; Length 448;  
Best Local Similarity 58.3%; Pred. No. 44;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 VEPVGSQHDS 12  
| | | | | : | : | :  
Db 274 VHTPSQKVDSQ 285

RESULT 23

US-09-272-796-2  
; Sequence 2, Application US/09272796  
; Patent No. 6207148  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Guegler, Karl G.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Goli, Surya K.  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN  
; TITLE OF INVENTION: KINASES  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/272,796  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/878,989  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0321 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 448 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: TBLYNOT01  
CLONE: 40194  
US-09-272-796-2

Query Match 59.7%; Score 37; DB 4; Length 448;  
Best Local Similarity 58.3%; Pred. No. 44;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 VEPVGSQHDSQ 12  
| | | | | : | : | :  
Db 274 VHTPSQKVDSQ 285

RESULT 24  
US-09-027-337-2  
; Sequence 2, Application US/09027337B  
; Patent No. 5972616  
; GENERAL INFORMATION:  
; APPLICANT: O'Brien, Timothy J.  
; APPLICANT: Tanimoto, Hirotschi  
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in  
; TITLE OF INVENTION: Breast and Ovarian Carcinomas  
; FILE REFERENCE: D6064  
; CURRENT APPLICATION NUMBER: US/09/027,337B

```
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of TADG-15 encoded by nucleotides
; OTHER INFORMATION: 23 to 2589 of Sequence 1
; Patent No. 5972616
US-09-027-337-2
```

```
Query Match      58.1%; Score 36; DB 2; Length 855;
Best Local Similarity 55.6%; Pred. No. 1.4e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 VEVPGSQHI 9
   :||| :||:
Db 370 LEVPPNQHV 378
```

```
RESULT 25
5194375-6
;Patent No. 5194375
; APPLICANT: PARK, LINDA S.;GOODWIN, RAYMOND G.
; TITLE OF INVENTION: DNA ENCODING INTERLEUKIN-7 RECEPTORS AND
;METHODS OF USE
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/493,588
; FILING DATE: 21-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 421,201
; FILING DATE: 13-OCT-1989
; APPLICATION NUMBER: 366,910
; FILING DATE: 15-JUN-1989
;SEQ ID NO:6
; LENGTH: 459
5194375-6
```

```
Query Match      54.8%; Score 34; DB 6; Length 459;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 VEVPGSQHI 9
   :|| ||||:
Db 46 LEVDGSQHL 54
```

Search completed: July 16, 2001, 16:36:18  
Job time: 207 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 16:37:00 ; Search time 34.9 Seconds  
(without alignments)  
26.192 Million cell updates/sec

Title: US-09-786-648-3

Perfect score: 62

Sequence: 1 VEPGSHDSQ 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62	100.0	124	1 XVVCB	cholera enterotoxin
2	62	100.0	124	1 QLECB	heat-labile entero
3	40	64.5	242	2 T34767	hypothetical prote
4	39	62.9	260	2 C96827	protein F20B17.2 I
5	39	62.9	272	2 T14755	hypothetical prote
6	39	62.9	353	2 H75446	(S)-2-hydroxy-acid
7	38	61.3	282	2 A33781	hippocampus-deri
8	38	61.3	674	1 A41670	carbon-monoxide de
9	37	59.7	352	2 B85518	hypothetical prote
10	37	59.7	367	2 T01751	gibberellin 20-oxi
11	37	59.7	376	1 S17246	chorismate synthas
12	37	59.7	423	2 G96554	hypothetical prote
13	37	59.7	442	2 T39683	zuotin-like protel
14	37	59.7	550	2 T01770	hypothetical prote
15	36	58.1	373	2 T47115	probable 4-carboxy
16	36	58.1	455	1 A69753	glucarate dehydrat
17	36	58.1	1785	2 S53976	probable membrane
18	35	56.5	142	2 S28757	cytochrome-c oxida
19	35	56.5	259	2 E84125	hydrolyase BH3805 I
20	35	56.5	284	2 F20331	methylentetrahidr
21	35	56.5	302	2 JN0143	catechol 1,2-dioxy
22	35	56.5	367	2 G83368	hypothetical prote
23	35	56.5	371	2 S23861	hypothetical prote
24	35	56.5	377	2 T23750	hypothetical prote
25	35	56.5	427	2 JC5694	stress-activated p
26	35	56.5	443	2 T23029	hypothetical prote
27	35	56.5	539	2 S53529	monophenol monooxy
28	35	56.5	593	2 T24379	hypothetical prote
29	35	56.5	670	2 T02092	beta-fructofuranos

30	35	56.5	960	2 A41638	chitin synthase (E
31	35	56.5	961	2 A53380	faclogonital dyspl
32	35	56.5	1258	2 T29041	hypothetical prote
33	35	56.5	2206	2 G71611	hypothetical prote
34	34	54.8	168	2 T27563	hypothetical prote
35	34	54.8	181	2 T45990	hypothetical prote
36	34	54.8	244	1 T01091	NADH dehydrogenase
37	34	54.8	252	2 B83584	probable molybdenu
38	34	54.8	269	2 T31648	hypothetical prote
39	34	54.8	274	2 G84353	hypothetical prote
40	34	54.8	285	2 T41734	hypothetical prote
41	34	54.8	295	2 C83388	transcriptional regu
42	34	54.8	365	2 G75126	cell division prot
43	34	54.8	365	2 A71005	probable cell divi
44	34	54.8	378	2 H82640	conserved hypothet
45	34	54.8	398	2 T15846	hypothetical prote

ALIGNMENTS

RESULT 1

XVVCB

cholera enterotoxin, B chain precursor VC1456 [validated] - Vibrio cholerae (strain N  
N:Alternate names: enterotoxin beta chain  
C:Species: Vibrio cholerae  
C:Date: 24-Apr-1984 #sequence\_revision 01-Sep-2000 #text\_change 02-Feb-2001  
C:Accession: S14624; S39241; H82196; JC1078; S17666; PC1010; A05130; A01819;  
R:Dams, E.; de Wolf, M.; Dierick, W.  
submitted to the EMBL Data Library, March 1991  
A:Description: Correction of the cholera toxin nucleotide sequence of the Vibrio chol  
A:Reference number: S14623  
A:Accession: S14624  
A:Molecule type: DNA  
A:Residues: 1-124 <DM>  
A:Cross-references: EMBL:X58786; NID:g48420; PIDN:CAA41593.1; PID:g48422  
A:Experimental source: strain 2125  
R:Lebens, M.; Holmgren, J.  
submitted to the EMBL Data Library, November 1993  
A:Description: Structure and arrangement of the Cholera toxin genes in vibrio Cholera  
A:Reference number: S39238  
A:Accession: S39238  
A:Molecule type: DNA  
A:Residues: 1-124 <LEB>  
A:Cross-references: EMBL:X76390; NID:g433856; PIDN:CAA53973.1; PID:g433857  
A:Accession: S39241  
A:Molecule type: DNA  
A:Residues: 1-124 <LEW>  
A:Cross-references: EMBL:X76391; NID:g433859; PIDN:CAA53976.1; PID:g433861  
R:Heidelber, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.  
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers  
I, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833  
A:Accession: H82196  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-124 <HEI>  
A:Cross-references: GB:AE004224; GB:AE003852; NID:g9655952; PIDN:AAF94613.1; GSPDB:GN  
R:Shi, C.H.; Cao, C.; Zhang, J.S.; Ma, Q.J.  
Chinese Biochem. J. 9, 395-399, 1993  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
A:Title: Nucleotide sequence analysis of the gene encoding the classical biotype V.ch  
A:Reference number: JC1078  
A:Accession: JC1078  
A:Molecule type: DNA  
A:Residues: 1-20, 'Q', 22-31, 'Q', 33-38, 'H', 40-49, 'G', 51-67, 'T', 69-124 <SHI>  
A:Experimental source: classical biotype strain 569B  
R:Dams, E.; de Wolf, M.; Dierick, W.  
Biochim. Biophys. Acta 1090, 139-141, 1991  
A:Title: Nucleotide sequence analysis of the CT operon of the Vibrio cholerae classic  
A:Reference number: S17665; MUID:91355224

A:Accession: S17666  
 A:Molecule type: DNA  
 A:Residues: 1-38, 'H', 40-67, 'T', 69-124 <DA2>  
 A:Cross-references: EMBL:X58785; NID:948888; PIDN:CAA1591.1; PID:g48890  
 R:Ma, Q.J.; Liu, C.X.; Xiong, L.S.; Yu, X.Q.  
 Sci. Sin. B Chem. Biol. Agric. Med. Earth Sci. 34, 274-280, 1991  
 A:Title: B subunit of cholera toxin produced in *Escherichia coli*.  
 A:Reference number: PC1010  
 A:Accession: PC1010  
 A:Molecule type: protein  
 A:Residues: 22-38, 'H', 40-41 <MAQ>  
 R:Mekalanos, J.J.; Swartz, D.J.; Pearson, G.D.N.; Harford, N.; Groyne, F.; de Wilde, M.  
 Nature 306, 551-557, 1983  
 A:Reference number: A93320; MUID:84068199  
 A:Accession: A05130  
 A:Molecule type: DNA  
 A:Residues: 1-32, 'S', 34-74, 'S', 76-124 <MEK>  
 A:Cross-references: GB:X00171; NID:g48347; PIDN:CAA24996.1; PID:g758351  
 R:Kurosky, A.; Markel, D.E.; Peterson, J.W.  
 J. Biol. Chem. 252, 7257-7264, 1977  
 A:Title: Covalent structure of the beta chain of cholera enterotoxin.  
 A:Reference number: A01819; MUID:78005537  
 A:Accession: A01819  
 A:Molecule type: protein  
 A:Residues: 22-38, 'H', 40-42, 'N', 44-67, 'T', 69-90, 'N', 92-124 <KUR>  
 R:Lai, C.Y.  
 J. Biol. Chem. 252, 7249-7256, 1977  
 A:Title: Determination of the primary structure of cholera toxin B subunit.  
 A:Reference number: A38033; MUID:78005536  
 A:Accession: A38033  
 A:Molecule type: protein  
 A:Residues: 22-38, 'H', 40-42, 'N', 44-67, 'T', 69, 'E', 71-90, 'N', 92-124 <LAI>  
 A:Note: the difference at residue 70 may be due to deamidation during preparation  
 R:Nakashima, Y.; Napiorkowski, P.; Schafer, D.E.; Konigsberg, W.H.  
 FEBS Lett. 68, 275-278, 1976  
 A:Title: Primary structure of the B subunit of cholera enterotoxin.  
 A:Reference number: A38034; MUID:77026365  
 A:Accession: A38034  
 A:Molecule type: protein  
 A:Residues: 22-38, 'H', 40-67, 'T', 69, 'E', 71, 'QS', 74-75, 'VE', 78-86, 'Q', 88-99, 'Q', 101-103, 'Q'  
 R:Rao, T.; Watanabe, H.; Shimonishi, Y.  
 Eur. J. Biochem. 146, 503-508, 1985  
 A:Title: Facile identification of protein sequences by mass spectrometry.  
 A:Reference number: A21910; MUID:85126976  
 A:Accession: A21910  
 A:Molecule type: protein  
 A:Residues: 22-38, 'H', 40-42, 'N', 44-67, 'T', 69-90, 'N', 92-124 <TAK>  
 A:Experimental source: biotype Inaba 569B  
 A:Note: Asn-65 was partially deaminated to Asp  
 C:Comment: The authors translated the codon TCA for residue 33 as Tyr.  
 C:Genetics:  
 A:Gene: VC1456  
 A:Map position: 1  
 C:Complex: the cholera enterotoxin molecule contains three kinds of chains; an alpha and  
 ciate noncovalently with the subunit B, an aggregate of five beta chains  
 C:Function:  
 C:Description: involved in binding of the toxin to cell membranes  
 C:Superfamily: cholera enterotoxin beta chain  
 C:Keywords: enterotoxin; toxin  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-124/Product: cholera enterotoxin chain B #status predicted <MAT>  
 F:30-107/Disulfide bonds: #status experimental

Query Match 100.0%; Score 62; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.00014;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12  
 |||||  
 Db 71 VEVPGSQHDSQ 82

RESULT 2  
 QLECB  
 heat-labile enterotoxin chain B precursor - *Escherichia coli*  
 C:Species: *Escherichia coli*  
 C:Date: 29-Jun-1981 #sequence\_revision 29-Jun-1981 #text\_change 18-Jun-1999  
 C:Accession: A01820; B26946; I41194; I41287; I67644; A61475  
 R:Dallas, W.S.; Falkow, S.  
 Nature 288, 499-501, 1980  
 A:Title: Amino acid sequence homology between cholera toxin and *Escherichia coli* heat  
 A:Reference number: A01820; MUID:81074965  
 A:Accession: A01820  
 A:Molecule type: mRNA  
 A:Residues: 1-124 <DAL>  
 R:Yamamoto, T.; Gojobori, T.; Yokota, T.  
 J. Bacteriol. 169, 1352-1357, 1987  
 A:Title: Evolutionary origin of pathogenic determinants in enterotoxigenic *Escherichia*  
 A:Reference number: A26946; MUID:87137303  
 A:Accession: B26946  
 A:Molecule type: DNA  
 A:Residues: 1-27, 'E', 29-63, 'K', 65-124 <YAM>  
 A:Cross-references: EMBL:M15363; NID:g148335; PIDN:AAA24792.1; PID:g148336  
 R:Leong, J.; Vinal, A.C.; Dallas, W.S.  
 Infect. Immun. 48, 73-77, 1985  
 A:Title: Nucleotide sequence comparison between heat-labile toxin B-subunit cistrons  
 A:Reference number: I41194; MUID:85156481  
 A:Accession: I41194  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-5, 'F', 7-17, 'C', 19-24, 'S', 26-27, 'E', 29-33, 'H', 35-63, 'K', 65-66, 'A', 68-122  
 A:Cross-references: GB:M17874; NID:g145830; PIDN:AAA98064.1; PID:g145831  
 R:Abrahami, I.; Gentz, R.  
 J. Biol. Chem. 262, 10189-10194, 1987  
 A:Title: A functional interaction between the signal peptide and the translation appa  
 ticulum.  
 A:Reference number: I41287; MUID:87280041  
 A:Accession: I41287  
 A:Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 1-22 <RE2>  
 A:Cross-references: GB:M17101; NID:g146375; PIDN:AAA23973.1; PID:g146376  
 R:Inoue, T.; Tsuji, T.; Koto, M.; Imamura, S.; Miyama, A.  
 FEBS Microbiol. Lett. 108, 157-161, 1993  
 A:Title: Amino acid sequence of heat-labile enterotoxin from chicken enterotoxigenic  
 A:Reference number: I53542; MUID:93252225  
 A:Accession: I67644  
 A:Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 1-17, 'C', 19, 'Y', 21-24, 'S', 26-27, 'E', 29-63, 'K', 65-66, 'A', 68-122, 'E', 124 <R  
 A:Cross-references: GB:S60731; NID:g408994; PIDN:AAC50441.1; PID:g408996  
 R:Tsuji, T.; Lida, T.; Honda, T.; Miwatani, T.; Nagahama, M.; Sakurai, J.; Wada, K.;  
 Microb. Pathog. 2, 381-390, 1987  
 A:Title: A unique amino acid sequence of the B subunit of a heat-labile enterotoxin 1  
 A:Reference number: A61475; MUID:89180953  
 A:Accession: A61475  
 A:Molecule type: protein  
 A:Residues: 22-24, 'S', 26-27, 'E', 29-63, 'K', 65-66, 'A', 68-95, 'A', 97-122, 'E', 124 <TSU>  
 A:Experimental source: strain 240-3  
 C:Complex: the heat-labile enterotoxin molecule contains one A chain and five or six  
 C:Function:  
 C:Description: the biological activity of the toxin is produced by the A chain, which  
 C:Superfamily: cholera enterotoxin beta chain  
 C:Keywords: enterotoxin  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-124/Product: heat-labile enterotoxin chain B #status predicted <MAT>  
 F:30-107/Disulfide bonds: #status predicted

Query Match 100.0%; Score 62; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.00014;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12

Db 71 VEPGSOHIDSQ 82  
|||||

## RESULT 3

hypothetical protein SC2A11.21c - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C>Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
C:Accession: T34767  
R:Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, August 1998  
A:Reference number: Z21556  
A:Accession: T34767  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-242 <MUR>  
A:Cross-references: EMBL:AL031184; PIDN:CAA20190.1; GSPDB:GN00070; SCOEDB:SC2A11.21c  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC2A11.21c

Query Match 64.5%; Score 40; DB 2; Length 242;  
Best Local Similarity 70.0%; Pred. No. 4.7;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VEPGSOHID 10  
|||||

Db 204 VEPGTDHTD 213

## RESULT 4

protein F20B17.2 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: C96827  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maithi, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: C96827  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-260 <STO>  
A:Cross-references: GB:AE005173; NID:g7715588; PIDN:AAF68106.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F20B17.2  
A:Map position: 1

Query Match 62.9%; Score 39; DB 2; Length 260;  
Best Local Similarity 50.0%; Pred. No. 7.9;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VEPGSOHIDSQ 12  
|||||

Db 65 VKVPGKHVSEK 76

## RESULT 5

hypothetical protein DKFp564A0122.1 - human  
C:Species: Homo sapiens (man)  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T14755  
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, August 1999  
A:Reference number: Z18181  
A:Accession: T14755  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-272 <WAM>  
A:Cross-references: EMBL:AL110209  
A:Experimental source: fetal brain; clone DKFp564A0122  
C:Genetics:  
A:Note: DKFp564A0122.1

Query Match 62.9%; Score 39; DB 2; Length 272;  
Best Local Similarity 66.7%; Pred. No. 8.3;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 EVPGSOHID 10  
|||||

Db 246 ELPGEHIE 254

## RESULT 6

H75446  
(S)-2-hydroxy-acid oxidase - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 28-Jul-2000  
C:Accession: H75446  
R:White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: H75446  
A:Molecule type: DNA  
A:Residues: 1-353 <WHI>  
A:Cross-references: GB:AE001954; GB:AE000513; NID:g6458751; PIDN:AAF10604.1; PID:g645  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR1031  
A:Map position: 1  
A:Superfamily: (S)-2-hydroxy-acid oxidase; (S)-2-hydroxy-acid oxidase homology  
F:3-297/Domain: (S)-2-hydroxy-acid oxidase homology <2HY>

Query Match 62.9%; Score 39; DB 2; Length 353;  
Best Local Similarity 75.0%; Pred. No. 11;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 VPGSOHID 10  
|||||

Db 189 VPGSEHLD 196

## RESULT 7

A35781  
hippocampus-derived neurotrophic factor precursor - rat  
N:Alternate names: neurotrophin-3 precursor  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 05-Oct-1990 #sequence\_revision 05-Oct-1990 #text\_change 16-Jul-1999  
C:Accession: A35781; A40094  
R:Ernfors, P.; Ibanez, C.F.; Ebendal, T.; Olson, L.; Persson, H.  
Proc. Natl. Acad. Sci. U.S.A. 87, 5454-5458, 1990  
A:Title: Molecular cloning and neurotrophic activities of a protein with structural s  
A:Reference number: A35781; MUID:90319130  
A:Accession: A35781  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-282 <ERN>  
A:Cross-references: GB:M34643

A:Accession: B95310  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-352 <STO>  
A:Cross-references: GB:AE005174; NID:g12513064; PIDN:AAG54606.1; GSPDB:GN00145; UWGP:Z03  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: 20347

Query Match 59.7%; Score 37; DB 1; Length 376;



Best Local Similarity 70.0%; Score 37; DB 2; Length 423;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VEPGSHQID 10  
| |||||  
Db 264 VSPGSKND 273

RESULT 12  
G96554  
hypothetical protein F19C24.16 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: G96554  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: G96554  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-423 <STO>  
A:Cross-References: GB:AE005173; NID:g11094753; PIDN:AAG29686.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F19C24.16  
A:Map position: 1

Query Match 59.7%; Score 37; DB 2; Length 423;  
Best Local Similarity 75.0%; Pred. No. 33;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VPGSHQID 10  
| |||||  
Db 403 VPGIHDV 410

RESULT 13  
T39683  
zootin-like protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C>Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 20-Oct-2000  
C:Accession: T39683; T40195  
R:Oliver, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, March 1998  
A:Reference number: Z21869  
A:Accession: T39683  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-124 <OLI>  
A:Cross-References: EMBL:AL049489; PIDN:CAB39796.1; GSPDB:GN00067; SPDB:SPBC1778.01c  
A:Experimental source: strain 972h-; cosmid c1778  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Lauber, J.; Hilbert, H.; Duesterhoeft, A.  
submitted to the EMBL Data Library, February 1998  
A:Reference number: Z21910  
A:Accession: T40195  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 89-442 <WOO>  
A:Cross-References: EMBL:Z97992; PIDN:CAB10796.1; GSPDB:GN00067; SPDB:SPBC30D10.01  
A:Experimental source: strain 972h-; cosmid c30D10  
C:Genetics:  
A:Gene: SPDB:SPBC1778.01c; SPDB:SPBC30D10.01  
A:Map position: 2

Query Match 59.7%; Score 37; DB 2; Length 442;  
Best Local Similarity 50.0%; Pred. No. 34;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVPGSHQIDS 11  
| |||||  
Db 371 DVPSAEHVD 380

RESULT 14  
T01770  
hypothetical protein A\_IG002P16.9 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 22-Oct-1999  
C:Accession: T01770  
R:Miller, N.; Beck, C.; Kramer, J.  
submitted to the EMBL Data Library, June 1997  
A:Description: The sequence of A. thaliana IG002P16.  
A:Reference number: Z14421  
A:Accession: T01770  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-550 <MIL>  
A:Cross-References: EMBL:AF007270; NID:g2191157; PID:g2191172; GSPDB:GN00063; ATSP:A\_IG002P16.9  
C:Genetics:  
A:Gene: ATSP:A\_IG002P16.9  
A:Map position: 5  
A:Introns: 159/1; 272/1; 434/2; 477/3

Query Match 59.7%; Score 37; DB 2; Length 550;  
Best Local Similarity 66.7%; Pred. No. 44;  
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VEPGSHQIDS 12  
| |||||  
Db 260 VTCGSGSHIDFQ 271

RESULT 15  
T47115  
probable 4-carboxymuconolactone decarboxylase / 3-oxoadipate enol-lactone hydrolase [  
C:Species: Streptomyces sp.  
C>Date: 17-Mar-2000 #sequence\_revision 17-Mar-2000 #text\_change 17-Mar-2000  
C:Accession: T47115  
R:Yang, K.; Iwagami, S.; Davies, J.E.  
submitted to the EMBL Data Library, May 1999  
A:Description: A protocatechuate catabolic gene cluster cloned from Streptomyces sp.  
A:Reference number: Z24354  
A:Accession: T47115  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-373 <YAN>  
A:Cross-References: EMBL:AF109386; PIDN:AAD40815.1  
A:Experimental source: strain 2065  
C:Genetics:  
A:Gene: pcal

Query Match 58.1%; Score 36; DB 2; Length 373;  
Best Local Similarity 55.6%; Pred. No. 44;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VEPGSHQI 9  
| |||||  
Db 225 VEIPGASHL 233

RESULT 16  
AG9753  
glucarate dehydratase (EC 4.2.1.40) - Bacillus subtilis  
C:Species: Bacillus subtilis  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000

C:Accession: A69753  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertea  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd  
A.; Ehrlich, S.D.; Emerson, P.T.; Enian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.  
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamane, H.; Yamamoto, K.; Yata, K.; Yoshida, K.  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033  
A:Accession: A69753  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-455 <KUN>  
A:Cross-references: GB:299105; GB:AL009126; NID:92632457; PIDN:CAB12043.1; PID:92632535  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: ycbF  
C:Superfamily: glucarate dehydratase  
C:Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 58.1%; Score 36; DB 1; Length 455;  
Best Local Similarity 75.0%; Pred. No. 55;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVPGSQHI 9  
||||:|  
Db 64 EVPGGEHI 71

RESULT 17  
S53976  
probable membrane protein YMR306w - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: hypothetical protein YW952.08  
C:Species: *Saccharomyces cerevisiae*  
C:Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 06-Feb-1998  
C:Accession: S53976  
R:Connor, R.; Churcher, C.M.  
submitted to the EMBL Data Library, April 1995  
A:Reference number: S53969  
A:Accession: S53976  
A:Molecule type: DNA  
A:Residues: 1-1785 <CON>  
A:Cross-references: EMBL:Z49212; NID:9798940; PID:9798948; MIPS:YMR306w  
C:Genetics:  
A:Gene: SGD:FKS3  
A:Cross-references: SGD:S0004923; MIPS:YMR306w  
A:Map position: 13R  
C:Keywords: transmembrane protein  
F:375-391/Domain: transmembrane #status predicted <TM1>  
F:415-431/Domain: transmembrane #status predicted <TM2>  
F:446-462/Domain: transmembrane #status predicted <TM3>  
F:560-576/Domain: transmembrane #status predicted <TM4>  
F:582-598/Domain: transmembrane #status predicted <TM5>  
F:1215-1231/Domain: transmembrane #status predicted <TM6>  
F:1268-1284/Domain: transmembrane #status predicted <TM7>  
F:1361-1377/Domain: transmembrane #status predicted <TM8>  
F:1385-1401/Domain: transmembrane #status predicted <TM9>  
F:1473-1489/Domain: transmembrane #status predicted <TM10>  
F:1507-1523/Domain: transmembrane #status predicted <TM11>  
F:1526-1542/Domain: transmembrane #status predicted <TM12>  
F:1551-1567/Domain: transmembrane #status predicted <TM13>  
F:1584-1600/Domain: transmembrane #status predicted <TM14>  
F:1655-1671/Domain: transmembrane #status predicted <TM15>  
F:1717-1733/Domain: transmembrane #status predicted <TM16>

C:Accession: A69753  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertea  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd  
A.; Ehrlich, S.D.; Emerson, P.T.; Enian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.  
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamane, H.; Yamamoto, K.; Yata, K.; Yoshida, K.  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033  
A:Accession: A69753  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-455 <KUN>  
A:Cross-references: GB:299105; GB:AL009126; NID:92632457; PIDN:CAB12043.1; PID:92632535  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: ycbF  
C:Superfamily: glucarate dehydratase  
C:Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 58.1%; Score 36; DB 2; Length 1785;  
Best Local Similarity 54.5%; Pred. No. 2.5e+02;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVPGSQHDSQ 12  
||||:|  
Db 403 EWFGAQLSSR 413

RESULT 18  
S28757  
cytochrome-c oxidase (EC 1.9.3.1) chain I - blue mussel mitochondrion (fragments)  
C:Species: mitochondrion *Mytilus edulis* (blue mussel)  
C:Date: 20-Feb-1995 #sequence\_revision 26-Jul-1996 #text\_change 07-Dec-1999  
C:Accession: S28757  
R:Hoffmann, R.J.; Boore, J.L.; Brown, W.M.  
Genetics 131, 397-412, 1992  
A:Title: A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*  
A:Reference number: S28743; MUID:92354892  
A:Accession: S28757  
A:Molecule type: DNA  
A:Residues: 1-43:44-142 <BRO>  
A:Cross-references: EMBL:M83761; EMBL:M83762  
C:Genetics:  
A:Gene: COI  
A:Genome: mitochondrion  
A:Genetic code: SGD  
C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology  
C:Keywords: copper; electron transfer; heme; magnesium; membrane-associated complex;

Query Match 56.5%; Score 35; DB 2; Length 142;  
Best Local Similarity 66.7%; Pred. No. 24;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 PGSQHDSQ 12  
||||:|  
Db 111 PGSKHVYSQ 119

RESULT 19  
E84125  
hydrolyase BH3805 [imported] - *Bacillus halodurans* (strain C-125)  
C:Species: *Bacillus halodurans*  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 31-Dec-2000  
C:Accession: E84125  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* a  
A:Reference number: A83650; MUID:20263314  
A:Accession: E84125  
A:Molecule type: DNA  
A:Status: preliminary  
A:Residues: 1-259 <STO>  
A:Cross-references: GB:AP001520; GB:BA000004; NID:g10176401; PIDN:BA07524.1; GSPDB:G  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH3805

Query Match 56.5%; Score 35; DB 2; Length 259;  
Best Local Similarity 60.0%; Pred. No. 46;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 VEVPGSQHID 10  
||||:|  
Db 135 VSVTGSSEHLD 144

RESULT 20

F82031  
methylenetetrahydrofolate dehydrogenase (NADP+) (EC 1.5.1.5) / methenyltetrahydrofolate  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: F82031  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
R:Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: A81775; MUID:20222556  
A:Accession: F82031  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-284 <PAR>  
A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83657.1; PID:g737910  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: f0LD; NMA0354  
C:Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate  
C:Keywords: hydrolase; oxidoreductase

Query Match 56.5%; Score 35; DB 2; Length 284;  
Best Local Similarity 50.0%; Pred. No. 51;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VEPGSHIDSQ 12  
|::| :|::|::|  
Db 97 VQLPLPKHDSQ 108

RESULT 21  
JN0143  
catechol 1,2-dioxygenase (EC 1.13.11.1) - Pseudomonas sp. plasmid EST1001  
C:Species: Pseudomonas sp.  
C:Date: 05-Mar-1993 #sequence\_revision 05-Mar-1993 #text\_change 29-Sep-1999  
R:Kivisaar, M.; Kasak, L.; Nurk, A.  
Gene 98, 15-20, 1991  
A:Title: Sequence of the plasmid-encoded catechol 1,2-dioxygenase-expressing gene, pheB,  
A:Reference number: JN0143; MUID:91192610  
A:Accession: JN0143  
A:Molecule type: DNA  
A:Residues: 1-302 <KIV>  
A:Cross-references: GB:M57500; NID:g145127; PIDN:AAC64900.1; PID:g145129  
C:Genetics:  
A:Gene: pheB  
A:Genome: plasmid  
C:Superfamily: catechol 1,2-dioxygenase  
C:Keywords: aromatic hydrocarbon catabolism; oxidoreductase

Query Match 56.5%; Score 35; DB 2; Length 302;  
Best Local Similarity 41.7%; Pred. No. 54;  
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 VEPGSHIDSQ 12  
:|::|::|  
Db 227 ISAPGHQHTTQ 238

RESULT 22  
G83368  
hypothetical protein PA2218 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: G83368  
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337

A:Accession: G83368  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-367 <STO>  
A:Cross-references: GB:AE004648; GB:AE004091; NID:g9948237; PIDN:AAG05606.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA2218

Query Match 56.5%; Score 35; DB 2; Length 367;  
Best Local Similarity 62.5%; Pred. No. 67;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VPGSQHID 10  
|::| :|::|  
Db 339 VPGASHVD 346

RESULT 23  
S23861  
hypothetical protein 2 - Pseudomonas aeruginosa  
C:Species: Pseudomonas aeruginosa  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 08-Oct-1999  
C:Accession: S23861  
R:Huang, H.; Siehnel, R.J.; Bellido, F.; Rawling, E.; Hancock, R.E.W.  
submitted to the EMBL Data Library, July 1992  
A:Description: Analysis of two gene regions involved in the expression of the imipene  
A:Reference number: S23859  
A:Accession: S23861  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-371 <HUA>  
A:Cross-references: EMBL:Z14064; NID:g45366; PIDN:CAA78447.1; PID:g45369

Query Match 56.5%; Score 35; DB 2; Length 371;  
Best Local Similarity 62.5%; Pred. No. 68;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VPGSQHID 10  
|::| :|::|  
Db 343 VPGASHVD 350

RESULT 24  
T29750  
hypothetical protein B0218.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000  
C:Accession: T29750  
R:Woesner, J.; Wohldmann, P.  
submitted to the EMBL Data Library, May 1996  
A:Description: The sequence of C. elegans cosmid B0218.  
A:Reference number: Z20678  
A:Accession: T29750  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-377 <WOE>  
A:Cross-references: EMBL:U58752; PIDN:AAB00564.1; GSPDB:GN000022; CESP:B0218.3  
A:Experimental source: strain Bristol N2; clone B0218  
C:Genetics:  
A:Gene: CESP:B0218.3  
A:Map position: 4  
A:Introns: 50/2; 239/1; 360/3  
C:Superfamily: kinase-related transforming protein; protein kinase homology

Query Match 56.5%; Score 35; DB 2; Length 377;  
Best Local Similarity 85.7%; Pred. No. 69;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PGSQHID 10

Db 235 PGSDHID 241  
||| |||

RESULT 25

JC5694  
stress-activated protein kinase (EC 2.7.-.-) JNRb - common carp  
N:Alternate names: stress-activated protein kinase b  
C:Species: Cyprinus carpio (common carp)  
C>Date: 20-Nov-1997 #sequence\_revision 20-Nov-1997 #text\_change 24-Sep-1999  
C:Accession: JC5694  
R:Hashimoto, H.; Matsuo, Y.; Yokoyama, Y.; Toyohara, H.; Sakaguchi, M.  
J. Biochem. 122, 381-386, 1997  
A:Title: Structure and expression of carp mitogen-activated protein kinases homologous to  
A:Reference number: JC5693; MUID:97456373  
A:Accession: JC5694  
A>Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-427 <HAS>  
A:Cross-references: DBJ:AB001744; NID:g2467307; PID:BAA22598.1; PID:d1023466; PID:g2467307  
C:Comment: This enzyme is a mitogen-activated protein kinase, and plays a role in egg maturation  
C:Superfamily: unassigned Ser/thr or tyr-specific protein kinases; protein kinase homologous to  
C:Keywords: phosphotransferase  
F:24-278/Domain: protein kinase homology <KIN>

Query Match 56.5%; Score 35; DB 2; Length 427;  
Best Local Similarity 85.7%; Pred. No. 79;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 4 PGSDHID 10  
||| |||  
Db 226 PGSDHID 232

Search completed: July 16, 2001, 16:37:01  
Job time: 210 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 16:44:03 ; Search time 19.88 Seconds  
(without alignments)  
20.677 Million cell updates/sec

Title: US-09-786-648-3  
Perfect score: 62  
Sequence: 1 VEPGSHIDSQ 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Swissprot\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	100.0	124	1	CHTB_VIBCH
2	62	100.0	124	1	ELBH_ECOLI
3	62	100.0	124	1	ELBP_ECOLI
4	38	61.3	674	1	DCMB_MOOTH
5	37	59.7	376	1	AROC_YEAST
6	36	58.1	455	1	GUDH_BACSU
7	36	58.1	743	1	BGAL_THEET
8	36	58.1	855	1	ST14_HUMAN
9	36	58.1	1785	1	GLS3_YEAST
10	35	56.5	302	1	PHB_PSESP
11	35	56.5	313	1	YP88_CAEEL
12	35	56.5	371	1	YM18_PSEAE
13	35	56.5	539	1	TYRO_ASPOR
14	35	56.5	670	1	INV1_MAIZE
15	35	56.5	960	1	CHS3_NEUCR
16	35	56.5	960	1	FGD1_MOUSE
17	35	56.5	961	1	FGD1_HUMAN
18	34	54.8	219	1	CIDA_HUMAN
19	34	54.8	244	1	NUHM_ARATH
20	34	54.8	365	1	FT23_PYRAB
21	34	54.8	365	1	FT23_PYRHO
22	34	54.8	419	1	GSC_DROME
23	34	54.8	432	1	AROC_NEUCR
24	34	54.8	456	1	SR54_THEAC
25	34	54.8	459	1	IL7R_MOUSE
26	34	54.8	491	1	CD5_RAT
27	34	54.8	500	1	GAST_HUMAN
28	34	54.8	508	1	Y202_HUMAN
29	34	54.8	557	1	HLVB_SERMA
30	34	54.8	560	1	INR1_SHEEP
31	34	54.8	775	1	TH1L_SCHPO
32	34	54.8	842	1	DPOL_HPBMV
33	33.5	54.0	375	1	MML7_MOUSE

34	33.5	54.0	375	1	MASP_RAT	P70564 rattus norv
35	33	53.2	259	1	MOB2_YEAST	P43563 saccharomyc
36	33	53.2	264	1	FTSQ_STRCU	O86038 streptomyce
37	33	53.2	386	1	YAAN_BACSU	P37535 bacillus su
38	33	53.2	454	1	VNUC_THOGV	P89216 thogoto vir
39	33	53.2	504	1	A37C_DROME	P18487 drosophila
40	33	53.2	534	1	YOG1_CAEEL	P34610 caenorhabdi
41	33	53.2	621	1	HEM1_AGABI	O92403 agaricus bi
42	33	53.2	666	1	PD14_MOUSE	O92183 mus musculu
43	33	53.2	773	1	MAK5_YEAST	P38112 saccharomyc
44	33	53.2	843	1	DPOL_HPBMV	P03157 hepatitis b
45	33	53.2	920	1	MML7_MYCTU	P96289 mycobacteri

ALIGNMENTS

RESULT 1	CHTB_VIBCH	STANDARD;	PRT;	124 AA.
ID	P01556; Q9JQ02;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	13-AUG-1987 (Rel. 05, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	CHOLERA ENTEROTOXIN, BETA CHAIN PRECURSOR.			
GN	CTXB OR TOXB OR VCL1456.			
OS	Vibrio cholerae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.			
OX	NCBI_TaxID=666;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=84061784; PubMed=6315707;			
RA	Lockman H., Kaper J.B.;			
RT	"Nucleotide sequence analysis of the A2 and B subunits of Vibrio cholerae enterotoxin.";			
RL	J. Biol. Chem. 258:13722-13726(1983).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=EL TOR 2125;			
RA	Mekalanos J.J., Swartz D.J., Pearson G.D.N., Harford N., Groyne F., de Wilde M.;			
RT	"Cholera toxin genes: nucleotide sequence, deletion analysis and vaccine development".			
RL	Nature 306:551-557(1983).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=EL TOR 2125;			
RA	Mekalanos J.J., Swartz D.J., Pearson G.D.N., Harford N., Groyne F., de Wilde M.;			
RT	"Cholera toxin genes: nucleotide sequence, deletion analysis and vaccine development".			
RL	Nature 306:551-557(1983).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=EL TOR 2125;			
RA	Dams E., de Wolf M., Dierick W.;			
RT	Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=4260B / SEROTYPE O139;			
RA	MEDLINE=94237453; PubMed=8181723;			
RL	Lebens M., Holmgren J.;			
RT	"Structure and arrangement of the cholera toxin genes in Vibrio cholerae O139.";			
RL	FEMS Microbiol. Lett. 117:197-202(1994).			
RN	[6]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=1854 / O139-BENGAL;			
RA	Yamamoto K., Do V.G.R.F., Xu M., Iida T., Miwatani T., Albert M.J., Honda T.;			
RT	Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.			
RN	[7]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=EL TOR N16961 / SEROTYPE O1;			
RA	MEDLINE=20406833; PubMed=10952301;			
RL	Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A., Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D., Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P., McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,			





RA MEDLINE=92114793; PUBMED=1837329;  
 RA Jones D.G.L., Reusser U., Braus G.H.;  
 RT "Molecular cloning, characterization and analysis of the regulation  
 RT of the ARO2 gene, encoding chorismate synthase, of *Saccharomyces*  
 RT *cerevisiae*.";   
 RL Mol. Microbiol. 5:2143-2152(1991).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=S288C / FY1769;  
 RP MEDLINE=97197983; PUBMED=9046099;  
 RX Voet M., Defoor E., Verhasselt P., Riles L., Robben J., Volckaert G.  
 RT "The sequence of a nearly unclonable 22.8 kb segment on the left arm  
 RT chromosome VII from *Saccharomyces cerevisiae* reveals ARO2, RPL9A,  
 RT TIPI, MRP1 genes and six new open reading frames.";   
 RL Yeast 13:177-182(1997).  
 RL -! CATALYTIC ACTIVITY: 5-O-((1-CARBOXYVINYL)-3-PHOSPHOSHIKIMATE -  
 CC CHORISMATE + ORTHOPHOSPHATE.  
 CC -! COFACTOR: REDUCED FLAVIN.  
 CC -! PATHWAY: SEVENTH STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN  
 CC THE BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY  
 CC -! SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -! INDUCTION: BY AMINO ACID STARVATION.  
 CC -! SIMILARITY: BELONGS TO THE CHORISMATE SYNTHASE FAMILY.



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X60190; CAA42745.1; -  
DR EMBL; X99960; CAA68214.1; -  
DR EMBL; Z72670; CAA96860.1; -  
DR PIR; S17246; S17246.  
DR SGD; S0003116; ARO2.  
DR InterPro; IPR000453; -  
DR Pfam; PF01264; Chorismate\_synt; 1.  
DR PROSITE; PS00787; CHORISMATE SYNTHASE\_1; 1.  
DR PROSITE; PS00788; CHORISMATE SYNTHASE\_2; 1.  
DR PROSITE; PS00789; CHORISMATE SYNTHASE\_3; 1.  
KW Lyase; Aromatic amino acid biosynthesis.  
SQ SEQUENCE 376 AA; 40838 MW; AF3AF65605B91E8E CRC64;  
  
Query Match 59.7%; Score 37; DB 1; Length 376;  
Best Local Similarity 70.0%; Pred. No. 14;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 VEVPQSQHID 10  
| | | | |  
DB 264 VSVPGSKHND 273  
  
RESULT 6  
GUDH\_BACSU STANDARD; PRT; 455 AA.  
AC P42238;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE PROBABLE GLUCARATE DEHYDRATASE (EC 4.2.1.40) (GDH) (GLUCD).  
GN GUDD.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95219079; PubMed=7704254;  
RA Ogawa K.-I., Akagawa E., Nakamura K., Yamane K.;  
RT "Determination of a 21548 bp nucleotide sequence around the 24  
degrees region of the Bacillus subtilis chromosome.";  
RL Microbiology 141:269-275(1995).  
CC -!- FUNCTION: CATALYZES THE DEHYDRATION OF GLUCARATE TO 5-KETO-4-  
DEOXY-D-GLUCARATE (5-KDGLUC) (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: D-GLUCARATE = 5-DEHYDRO-4-DEOXY-D-GLUCARATE +  
H(2)O.  
CC  
CC -!- PATHWAY: FIRST STEP IN GLUCARATE CATABOLISM.  
CC -!- SIMILARITY: BELONGS TO THE MANDELATE RACEMASE / MUONATE  
LACTONIZING ENZYME FAMILY. GLUCD SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D30808; BAA06470.1; -  
DR EMBL; Z99105; CAB12043.1; -  
DR Subtilist; BG11161; gudd.  
KW Lyase.

SQ SEQUENCE 455 AA; 50782 MW; 3238486007698C2A CRC64;  
  
Query Match 58.1%; Score 36; DB 1; Length 455;  
Best Local Similarity 75.0%; Pred. No. 26;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 2 EYVPGSQHI 9  
| | | | |  
DB 64 EYVPGGEHI 71  
  
RESULT 7  
BGAL\_THEET STANDARD; PRT; 743 AA.  
ID BGAL\_THEET  
AC P77989;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE BETA-GALACTOSIDASE (EC 3.2.1.23) (LACTASE).  
GN LACZ OR LACA.  
OS Thermoanaerobacter ethanolicus (Clostridium thermohydrosulfuricum).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Thermoanaerobacter group; Thermoanaerobacter.  
OX NCBI\_TaxID=1757;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33223 / 39E;  
RA Zverlov V.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-  
GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.  
CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF GLYCOSYL HYDROLASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Y08557; CAA69850.1; -  
DR InterPro; IPR001649; -  
DR Pfam; PF00703; Glyco\_hydro\_2; 1.  
DR PROSITE; PS00719; GLYCOSYL\_HYDROL\_F2\_1; 1.  
DR PROSITE; PS00608; GLYCOSYL\_HYDROL\_F2\_2; 1.  
KW Hydrolase; Glycosidase.  
FT ACT\_SITE 388 388 PROTON DONOR (BY SIMILARITY).  
FT ACT\_SITE 453 453 NUCLEOPHILE (BY SIMILARITY).  
SQ SEQUENCE 743 AA; 85796 MW; FE011FF517E51DFC CRC64;  
  
Query Match 58.1%; Score 36; DB 1; Length 743;  
Best Local Similarity 54.5%; Pred. No. 45;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 EYVPGSQHDSQ 12  
| | | | |  
DB 346 EYVPGWHIGDE 356  
  
RESULT 8  
STI4\_HUMAN STANDARD; PRT; 855 AA.  
ID STI4\_HUMAN  
AC Q9Y5Y6;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE SUPPRESSOR OF TUMORIGENICITY 14 (EC 3.4.21.-) (MATRIPTASE) (MEMBRANE-  
DE TYPE SERINE PROTEASE 1) (MT-SP1).  
GN STI4 OR PRS14 OR SNCL9.  
OS Homo sapiens (Human).



ID	YP68_CAEEL	STANDARD;	PRT;	313 AA.
Q09217;				
DT	01-NOV-1997	(Rel. 35, Created)		
DT	01-NOV-1997	(Rel. 35, Last sequence update)		
DT	01-OCT-2000	(Rel. 40, Last annotation update)		
DE	HYPOTHETICAL 37.0 KDA PROTEIN B0495.8 IN CHROMOSOME II.			
GN	B0495.8.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;			
OC	Rhabditidae; Pelodierinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BRISTOL N2;			
RA	Kirsten J.;			
RL	Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.			
CC	- !- SIMILARITY: SOME, TO YEAST YDL087C AND S.POMBE SPCC16A11.13.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; U21317; AAA62527.1; -			
DR	WormPep; B0495.8; CE01766.			
KW	Hypothetical protein			
SEQUENCE	313 AA; 36977 MW; 000D2327621BFED0	CRC64;		
QY	2 EYPSQHDSQ 12	Query Match	56.5%;	Score 35; DB 1; Length 313;
	:: :::: :::	Best Local Similarity	45.5%;	Pred. No. 27;
DB	11 QLMGSHVDNK 21	Matches	5; Conservative	5; Mismatches 1; Indels 0; Gaps 0;
RESULT 12				
YML18_PSEAE		STANDARD;	PRT;	371 AA.
ID	YML18_PSEAE			
AC	Q01609; Q911P9;			
DT	01-FEB-1995	(Rel. 31, Created)		
DT	01-OCT-2000	(Rel. 40, Last sequence update)		
DT	01-OCT-2000	(Rel. 40, Last annotation update)		
DE	HYPOTHETICAL PROTEIN PA2218.			
GN	PA2218.			
OS	Pseudomonas aeruginosa.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;			
OC	Pseudomonas.			
OX	NCBI_TaxID=287;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 15692 / PAO1 / H103;			
RC	MEDLINE=93051258; PubMed=1427017;			
RA	Huang H., Siehnell R.J., Bellido F., Rawling E., Hancock R.E.W.;			
RT	"Analysis of two gene regions involved in the expression of the			
RT	imipenem-specific, outer membrane porin protein Oprd of Pseudomonas			
RT	aeruginosa."			
RL	FEMS Microbiol. Lett. 76:267-274(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PAO1;			
RC	MEDLINE=20437337; PubMed=10984043;			
RA	Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,			
RA	Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,			
RA	Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,			
RA	Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,			
RA	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,			
RA	Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;			

RA Brody L.L., Coulter S.N., Folger K.K., Kas A., Larbig K., Lim K.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Saiter M.H., Hancock R.E.W., Lory S., Olson M.V.,  
RA Reizer J., Saiter M.H., Hancock R.E.W., Lory S., Olson M.V.,

RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an  
RT opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
CC -1- SIMILARITY: TO E.COLI YCJY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z14064; CRA78447.1; .  
DR EMBL; AF004648; AAG05606.1; ALT\_INIT.  
DR PIR; S23861; S23861.  
KW Hypothetical protein.  
FT CONFLICT 16 16 R -> P (IN REF. 1).  
FT CONFLICT 73 73 R -> P (IN REF. 1).  
FT CONFLICT 261 261 A -> G (IN REF. 1).  
SQ SEQUENCE 371 AA; 40840 MW; D7EB0CCAC95A7CF6 CRC64;  
  
Query Match 56.5%; Score 35; DB 1; Length 371;  
Best Local Similarity 62.5%; Pred. No. 32;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 3 VPGSQHID 10  
| | | : | |  
DB 343 VPGASHVD 350  
  
RESULT 13  
TYRO\_ASOR  
ID TYRO\_ASOR STANDARD; PRT; 539 AA.  
AC Q00234;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE TYROSINASE (EC 1.14.18.1) (MONOPHENOL MONOOXYGENASE).  
GN MELO.  
OS Aspergillus oryzae.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=5062;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=ATCC 22788 / RIB 128;  
RX MEDLINE=95200965; PubMed=7893753;  
RA Fujita Y., Uraga Y., Ichishima E.;  
RT "Molecular cloning and nucleotide sequence of the protyrosinase gene,  
RT melo, from Aspergillus oryzae and expression of the gene in yeast  
RT cells.";  
RL Biochim. Biophys. Acta 1261:151-154(1995).  
CC -1- FUNCTION: THIS IS A COPPER-CONTAINING OXIDASE THAT FUNCTIONS IN  
CC THE FORMATION OF PIGMENTS SUCH AS MELANINS AND OTHER POLYPHENOLIC  
CC COMPOUNDS.  
CC -1- CATALYTIC ACTIVITY: L-TYROSINE + L-DOPA + O(2) -> L-DOPA +  
CC DOPAQUINONE + H(2)O.  
CC -1- COFACTOR: BINDS TWO COPPER IONS.  
CC -1- ENZYME REGULATION: ACTIVATED BY ACIDIFYING TREATMENT AT PH 3.0.  
CC -1- SUBUNIT: HOMOTETRAMER.  
CC -1- PTM: THE N-TERMINAL IS BLOCKED.  
CC -1- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; D37929; BAA07149.1; .  
DR InterPro; IPR002227; .  
DR Pfam; PF00264; tyrosinase; 2.  
DR PRINTS; PR00092; TYROSINASE.  
DR PROSITE; PS00497; TYROSINASE\_1; 1.  
DR PROSITE; PS00498; TYROSINASE\_2; 1.  
KW Melanin biosynthesis; Oxidoreductase; Monooxygenase; Copper.  
FT METAL 63 63 COPPER A (BY SIMILARITY).  
FT METAL 84 84 COPPER A (BY SIMILARITY).  
FT METAL 93 93 COPPER B (BY SIMILARITY).  
FT METAL 290 290 COPPER B (BY SIMILARITY).  
FT METAL 294 294 COPPER B (BY SIMILARITY).  
FT METAL 333 333 COPPER B (BY SIMILARITY).  
FT THIOPTH 82 84 BY SIMILARITY.  
SQ SEQUENCE 539 AA; 60604 MW; CD2ECD702A018E15 CRC64;  
  
Query Match 56.5%; Score 35; DB 1; Length 539;  
Best Local Similarity 54.5%; Pred. No. 49;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 EYFGSHIDSQ 12  
| | | : | | |  
DB 360 EYFGTNSVDSQ 370  
  
RESULT 14  
INVL\_MAIZE  
ID INVL\_MAIZE STANDARD; PRT; 670 AA.  
AC P49175;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE BETA-FRUCTOFURANOSIDASE 1 PRECURSOR (EC 3.2.1.26) (SUCROSE-6-PHOSPHATE  
DE HYDROLASE 1) (INVERTASE 1).  
GN IVRI.  
OS Zea mays (Maize).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae;  
OC Andropogoneae; Zea.  
OX NCBI\_TaxID=4577;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. B73;  
RX MEDLINE=95357417; PubMed=7630946;  
RA Xu J., Pemberton G.H., Almira E.C., McCarty D.R., Koch K.E.;  
RT "The ivr 1 gene for invertase in maize.";  
RL Plant Physiol. 108:1293-1294(1995).  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL NON-REDUCING BETA-D-  
CC FRUCTOFURANOSIDE RESIDUES IN BETA-D-FRUCTOFURANOSIDES.  
CC -1- SUBCELLULAR LOCATION: VACUOLAR.  
CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U16123; AAA83439.1; .  
DR MaizEDB; 86037; .  
DR InterPro; IPR001362; .  
DR Pfam; PF00251; Glyco\_hydro\_32; 1.  
DR PROSITE; PS00609; GLYCOSYL\_HYDROL\_F32; 1.  
KW Hydrolase; Glycosidase; Glycoprotein; Zymogen; Signal;  
KW Multigene family.  
FT SIGNAL 1 ?  
FT PROPEP ? ?  
FT CHAIN ? 670  
FT ACT\_SITE 139 139 BETA-FRUCTOFURANOSIDASE 1.  
FT CARBOHYD 165 165 BY SIMILARITY.  
FT N-LINKED (GLCNAC. .) (POTENTIAL).

```

FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 518 518 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 595 595 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 670 AA; 71932 MW; DEDE0989C7E6AEB0 CRC64;

Query Match 56.5%; Score 35; DB 1; Length 670;
Best Local Similarity 58.3%; Pred. No. 62;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12
   |||||
Db 61 VYVLSQHDGQ 72

RESULT 15
ID CHS3_NEUCR STANDARD; PRT; 960 AA.
AC P29070;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE CHITIN SYNTHASE 3 (EC 2.4.1.16) (CHITIN-UDP ACETYL-GLUCOSAMINYL
DE TRANSFERASE 3) (CLASS-III CHITIN SYNTHASE 3).
GN CHS-3.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92090722; PubMed=1836444;
RA Farden O., Yanofsky C.;
RT "Chitin synthase 1 plays a major role in cell wall biogenesis in
RT Neurospora crassa.";
RL Genes Dev. 5:2420-2430(1991).
CC !- FUNCTION: PLAYS A MAJOR ROLE IN CELL WALL BIOGENESIS.
CC !- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + [1,4-(N-ACETYL-
CC BETA-D-GLUCOSAMINYL)](N) -> UDP + [1,4-(N-ACETYL-BETA-D-
CC GLUCOSAMINYL)](N+1).
CC !- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PLASMA MEMBRANE.
CC !- SIMILARITY: BELONGS TO THE CHITIN SYNTHASE FAMILY. SUBFAMILY CLASS
CC III.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M73437; AAA3568.1; -.
CC DR PIR; A41638; A41638.
CC DR InterPro; IPR002923; -.
CC DR Pfam; PF01644; Chitin_synth; 1.
CC KW Transferase; Glycosyltransferase; Transmembrane; Cell wall;
CC Multigene family.
CC FT TRANSMEM 571 591 POTENTIAL.
CC FT TRANSMEM 630 650 POTENTIAL.
CC FT TRANSMEM 665 685 POTENTIAL.
CC FT TRANSMEM 717 737 POTENTIAL.
CC FT TRANSMEM 745 765 POTENTIAL.
CC FT TRANSMEM 845 867 POTENTIAL.
CC SQ SEQUENCE 960 AA; 106903 MW; 633A2107319BF447 CRC64;

Query Match 56.5%; Score 35; DB 1; Length 960;
Best Local Similarity 66.7%; Pred. No. 92;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 PGSQHDSQ 12

```

```

Db 24 PGQHDHDAQ 32
||| ||| |||

FGDL_MOUSE
ID FGDL_MOUSE STANDARD; PRT; 960 AA.
AC P52734;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PUTATIVE RHO/RAC GUANINE NUCLEOTIDE EXCHANGE FACTOR (RHO/RAC GEF)
DE (FACIOGENITAL DYSPLASIA PROTEIN HOMOLOG).
GN FGDL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96081343; PubMed=8535076;
RA Pastoris N.G., de Gouyon B., Cadle A.B., Campbell K., Herman G.E.,
RA Gorski J.L.;
RT "Cloning and regional localization of the mouse facio-genital
RT dysplasia (Fgdl) gene.";
RL Mamm. Genome 6:658-661(1995).
CC !- FUNCTION: ACTIVATES THE RAS-LIKE FAMILY OF RHO- AND RAC PROTEINS
CC BY EXCHANGING BOUND GDP FOR FREE GTP.
CC !- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC !- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC !- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U22325; AAA96001.1; -.
CC DR MGD; MGI:104566; Fgdl.
CC DR InterPro; IPR000219; -.
CC DR InterPro; IPR000306; -.
CC DR InterPro; IPR000822; -.
CC DR InterPro; IPR001849; -.
CC DR Pfam; PF01363; FYVE; 1.
CC DR Pfam; PF00169; PH; 2.
CC DR Pfam; PF00621; RhoGEF; 1.
CC DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
CC DR PROSITE; PS00003; PH_DOMAIN; 2.
CC KW Guanine-nucleotide releasing factor; Zinc-finger.
CC DOMAIN 7 330 PRO-RICH.
CC FT DOMAIN 171 179 SH3-BINDING (POTENTIAL).
CC FT DOMAIN 179 187 SH3-BINDING (POTENTIAL).
CC FT DOMAIN 589 688 PH.
CC FT ZN_FING 733 757 C2H2-TYPE.
CC FT DOMAIN 820 920 PH.
CC SQ SEQUENCE 960 AA; 106477 MW; 41C1B84DE490FC51 CRC64;

Query Match 56.5%; Score 35; DB 1; Length 960;
Best Local Similarity 75.0%; Pred. No. 92;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 PGSOHIDS 11
   |||||
Db 330 PGSQEVDSD 337

RESULT 17
ID FGDL_HUMAN STANDARD; PRT; 961 AA.

```

P98174;  
01-OCT-1996 (Rel. 34, Created)  
01-OCT-1996 (Rel. 34, Last sequence update)  
01-OCT-2000 (Rel. 40, Last annotation update)  
01-OCT-2000 (Rel. 40, Last annotation update)  
DE PUTATIVE RHO/RAC GUANINE NUCLEOTIDE EXCHANGE FACTOR (RHO/RAC GEF)  
DE (FACIOGENITAL DYSPLASIA PROTEIN).  
GN FGDI.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Craniofacial;  
RX MEDLINE=95042764; PubMed=7954831;  
RA Pasteris N.G., Cadle A., Logie L.J., Porteous M.E.M., Schwartz C.E.,  
RT Stevenson R.E., Glover T.W., Wilroy R.S., Gorski J.L.;  
RT "Isolation and characterization of the facio-genital dysplasia  
RT (Aarskog-Scott syndrome) gene: a putative Rho/Rac guanine nucleotide  
RT exchange factor";  
RL Cell 79:669-678(1994).  
CC -!- FUNCTION: ACTIVATES THE RAS-LIKE FAMILY OF RHO- AND RAC PROTEINS  
CC -!- BY EXCHANGING BOUND GDP FOR FREE GTP.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
CC -!- TISSUE SPECIFICITY: FETAL HEART, BRAIN, LUNG, KIDNEY, PLACENTA,  
CC AND LESS IN LIVER; ADULT HEART, BRAIN, LUNG, SKELETAL MUSCLE, AND  
CC LESS IN PANCREAS AND LIVER.  
CC -!- DISEASE: FACIOGENITAL DYSPLASIA (AARSKOG-SCOTT SYNDROME) IS A RARE  
CC MULTISYSTEMIC DISORDER CHARACTERIZED BY DISPROPORTIONATELY SHORT  
CC STATURE, AND BY FACIAL, SKELETAL, AND UROGENITAL ANOMALIES.  
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).  
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U11690; AAA57004.1; -.  
CC MIM; 305400; -.  
CC InterPro; IPR000219; -.  
CC InterPro; IPR000306; -.  
CC InterPro; IPR000822; -.  
CC InterPro; IPR001849; -.  
CC Pfam; PF01363; FYVE; 1.  
CC Pfam; PF00169; PH; 2.  
CC Pfam; PF00621; RhoGEF; 1.  
CC PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
CC PROSITE; PS50003; PH\_DOMAIN; 2.  
KW Guanine-nucleotide releasing factor; Zinc-finger.  
FT DOMAIN 7 330 PRO-RICH.  
FT DOMAIN 171 179 SH3-BINDING (POTENTIAL).  
FT DOMAIN 179 187 SH3-BINDING (POTENTIAL).  
FT DOMAIN 590 689 PH.  
FT ZN\_FING 734 758 C2H2-TYPE.  
FT DOMAIN 821 921 PH.  
SQ SEQUENCE 961 AA; 106668 MW; CF722598853A685A CRC64;

Query Match 56.5%; Score 35; DB 1; Length 961;  
Best Local Similarity 75.0%; Pred. No. 92;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 PGSQHDS 11

Db 330 PGSQEVDS 337

RESULT 18

CIDA\_HUMAN

ID CIDA\_HUMAN STANDARD; PRT; 219 AA.  
AC O60543;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE CELL DEATH ACTIVATOR CIDE-A (CELL DEATH-INDUCING DIFFA-LIKE EFFECTOR  
DE A).  
DE GN CIDEA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98232498; PubMed=9564035;  
RA Inohara N., Koseki T., Chen S., Wu X., Nunez G.;  
RT "CIDE, a novel family of cell death activators with homology to the 45  
RT kDa subunit of the DNA fragmentation factor";  
RL EMBO J. 17:2526-2533(1998).  
CC -!- FUNCTION: ACTIVATES APOPTOSIS.  
CC -!- SUBUNIT: INHIBITED BY DEFB.  
CC -!- SIMILARITY: CONTAINS 1 CIDE-N DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF041378; AAC34987.1; -.  
CC MIM; 604440; -.  
CC Pfam; PF02017; CIDE-N; 1.  
KW Apoptosis.  
FT DOMAIN 33 110 CIDE-N.  
SQ SEQUENCE 219 AA; 24686 MW; 05F704823CE71C0E CRC64;

Query Match 54.8%; Score 34; DB 1; Length 219;  
Best Local Similarity 71.4%; Pred. No. 28;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 VPGSQHI 9  
Db 109 MPGSQHV 115

RESULT 19  
NUHM\_ARATH  
ID NUHM\_ARATH STANDARD; PRT; 244 AA.  
AC O22769;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE NADH-UBIQUINONE OXIDOREDUCTASE 24 KDA SUBUNIT PRECURSOR (EC 1.6.5.3)  
DE (EC 1.6.99.3).  
GN AT4G02580 OR T10P11.14.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=20083488; PubMed=10617198;  
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,  
RA Pohl T., Dueterhoeft A., Stiekema W., Entian K.-D., Terryn N.,  
RA Harris B., Ansonge W., Brandt P., Grivell L., Rieger M.,  
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,  
RA Kreis M., Delisny M., Puigdomenech P., Watson M., Schmidheini T.,  
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,

RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,  
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,  
RA Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandenbussche F.,  
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,  
RA Weitzenecker T., Bothe G., Ransperger U., Hilbert H., Braun M.,  
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirks W.,  
RA Mooljman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,  
RA Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,  
RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,  
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,  
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,  
RA Pettett A., Rajandream M.-A., Lyne M., Benes V., Rechmann S.,  
RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,  
RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,  
RA Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,  
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,  
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,  
RA Chedof F., Cooke R., Berger C., Monfort A., Casacuberta E.,  
RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,  
RA Perez-Perez A., Fumelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,  
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
RA Zaccarini P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,  
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,  
RA Sehkon M., Murray J., Sheet P., Cordes M., Abu-Threiden J.,  
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp C.,  
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,  
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,  
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,  
RA Antonolu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,  
RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,  
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,  
RA Chen E., Warra M., Martienssen R., McCombie W.R.,  
RA "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
thaliana";  
RT Nature 402:769-777(1999).  
RL -----  
CC !- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY  
CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED  
CC TO BE UBIQUINONE. COMPONENT OF THE FLAVOPROTEIN-SULFUR (FP)  
CC FRAGMENT OF THE ENZYME (BY SIMILARITY).  
CC !- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.  
CC !- COFACTOR: 2FE-2S IRON-SULFUR CLUSTER N 1B (BY SIMILARITY).  
CC !- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 40 DIFFERENT SUBUNITS.  
CC THIS IS A COMPONENT OF THE FLAVOPROTEIN FRACTION.  
CC !- SUBCELLULAR LOCATION: MATRIX AND CYTOPLASMIC SIDE OF THE  
CC MITOCHONDRIAL INNER MEMBRANE (BY SIMILARITY).  
CC !- SIMILARITY: BELONGS TO THE COMPLEX I 24 KDA SUBUNIT FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AC002330; AAC78260.1; -;  
CC EMBL; AL161494; CAB80751.1; -;  
CC InterPro; IPR002023; -;  
CC Pfam; PF01257; complex1\_24kdc; 1;  
CC PROSITE; PS01099; COMPLEX1\_24K; 1;  
CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transit peptide;  
KW Iron-sulfur. 1 ?  
FT TRANSIT ? 244  
FT CHAIN ?  
FT METAL 119 119  
FT METAL 124 124  
FT METAL 160 160  
FT METAL 160 160

FT METAL 164 164  
SQ SEQUENCE 244 AA; 27182 MW; 27C95BF5884B12AC CRC64;  
  
Query Match 54.8%; Score 34; DB 1; Length 244;  
Best Local Similarity 75.0%; Pred. No. 31;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 4 PGSQHIDS 11  
DB 20 PTSQHILDS 27  
  
RESULT 20  
FTZ3\_PYRAB STANDARD; PRT; 365 AA.  
AC Q9V0H5;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE CELL DIVISION PROTEIN FTSZ HOMOLOG 3.  
GN FTSZ3 OR PAB1820.  
OS Pyrococcus abyssii.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
OX NCBI\_TaxID=29292;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ORSAY;  
RA Heilig R.;  
RT "Pyrococcus abyssii genome sequence: Insights into archaeal chromosome  
RT structure and evolution.";  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AJ248285; CAB49728.1; -;  
CC InterPro; IPR000158; -;  
CC InterPro; IPR000217; -;  
CC PRINTS; PR00423; CELLDVIFTSZ.  
CC PRINTS; PR01161; TUBULIN.  
CC PROSITE; PS00227; TUBULIN; 1.  
CC Cell division; Septation; GTP-binding; Multigene family.  
FT NP\_BIND 99 107 GTP (POTENTIAL).  
SQ SEQUENCE 365 AA; 39957 MW; ACC31761552F0CC5 CRC64;  
  
Query Match 54.8%; Score 34; DB 1; Length 365;  
Best Local Similarity 45.5%; Pred. No. 49;  
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 VEVPQSQHIDS 11  
DB 195 IDVPGQTLDA 205  
  
RESULT 21  
FTZ3\_PYRHO STANDARD; PRT; 365 AA.  
AC O59060.  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE CELL DIVISION PROTEIN FTSZ HOMOLOG 3.  
GN FTSZ3 OR PH1335.  
OS Pyrococcus horikoshii.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
OX NCBI\_TaxID=53953;



RA SEQUENCE FROM N.A.  
RC STRAIN=OT3;  
RX MEDLINE=98344137; PubMed=9679194;  
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,  
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
RA Funahashi T., Tanaka T., Kudo H., Yamazaki J., Kishida N., Oguchi A.,  
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
RA Masuchi Y., Shizuya H., Kikuchi H.;  
RT "Complete sequence and gene organization of the genome of a hyper-  
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
RL DNA Res. 5:55-76(1998).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AP000006; BAA30441.1; -;  
DR InterPro; IPR000217; -;  
DR InterPro; IPR003008; -;  
DR Pfam; PF00091; tubulin; 1.  
DR PROSITE; PS00227; TUBULIN; 1.  
KW Cell division; Septation; GTP-binding; Multigene family.  
FT NP\_BIND 99 107  
FT SEQUENCE 365 AA; 40118 MW; 6C987E91C761F5B2 CRC64;  
SQ  
Query Match 54.8%; Score 34; DB 1; Length 365;  
Best Local Similarity 45.5%; Pred. No. 49;  
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
QY 1 VEVPGSQHDS 11  
Db 195 IDVPGEQTLDA 205  
::||| | :|  
195 IDVPGEQTLDA 205  
RESULT 22  
GSC\_DROME  
ID GSC\_DROME STANDARD; PRT; 419 AA.  
AC P54366;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE HOMEOBOX PROTEIN GOOSECOID.  
GN GSC.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=96202483; PubMed=8625850;  
RA Goriely A., Stella M., Coffinier C., Kessler D., Mailhos C.,  
RA Dessain S., Desplan C.;  
RT "A functional homologue of gooseoid in Drosophila.";  
RL Development 122:1641-1650(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96272167; PubMed=8670808;  
RA Hahn M., Jackle H.;  
RT "Drosophila gooseoid participates in neural development but not in  
RT body axis formation.";  
RL EMBL J. 15:3077-3084(1996).  
CC -1- FUNCTION: APPEARS TO REGULATE REGIONAL DEVELOPMENT OF SPECIFIC  
CC TISSUES. CAN RESCUE AXIS POLARITY IN UV-RADIATED XENOPUS  
CC EMBRYOS.

CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- TISSUE SPECIFICITY: IN EARLY EMBRYO DEVELOPMENT, EXPRESSION  
CC CONFINED TO TWO REGIONS: A HORSESHOE-LIKE PATTERN ACROSS THE  
CC DORSAL SIDE WHICH IS DESTINED TO FORM THE BRAIN HEMISPHERES AND A  
CC SECOND DOMAIN WHICH INVAGINATES INSIDE THE STOMODEUM AND WHICH, IS  
CC FATED TO FORM THE FOREGUT, RING GLAND AND STOMATOGASTRIC NERVOUS  
CC SYSTEM (SNS).  
CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEOBOX PROTEINS.  
CC "BICOID" SUBFAMILY.  
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-5 IS THE INITIATOR.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X95420; CAA64699.1; -;  
DR EMBL; U52968; AAB17948.1; -;  
DR HSSP; P06601; 1FJL.; -;  
DR FlyBase; FBgn0010323; Gsc.  
DR InterPro; IPR001356; -;  
DR Pfam; PF00046; homeobox; 1.  
DR PROSITE; PS00027; HOMEOBOX\_1; 1.  
DR PROSITE; PS00071; HOMEOBOX\_2; 1.  
KW Developmental protein; Nuclear protein; DNA-binding; Homeobox.  
FT DOMAIN 104 107 POLY-ALA.  
FT DOMAIN 164 169 POLY-SER.  
FT DOMAIN 195 199 POLY-ALA.  
FT DNA\_BIND 286 345 HOMEOBOX.  
FT SEQUENCE 419 AA; 44949 MW; 851A4C46AA861FB9 CRC64;  
SQ  
Query Match 54.8%; Score 34; DB 1; Length 419;  
Best Local Similarity 50.0%; Pred. No. 57;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 1 VEVPGSQHDSQ 12  
Db 385 VKCPGSDHSAQ 396  
::||| | :|  
12 VEVPGSQHDSQ 12  
385 VKCPGSDHSAQ 396  
RESULT 23  
AROC\_NEUCR  
ID AROC\_NEUCR STANDARD; PRT; 432 AA.  
AC Q12640;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE CHORISMATE SYNTHASE (EC 4.6.1.4) (5-ENOLPYRUVYL-SHIMADZU-3-PHOSPHATE  
DE PHOSPHOLYASE).  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95386486; PubMed=7657620;  
RA Henstrand J.M., Amrhein N., Schmid J.;  
RT "Cloning and characterization of a heterologously expressed  
RT bifunctional chorismate synthase/flavin reductase from Neurospora  
RT crassa.";  
RL J. Biol. Chem. 270:2047-2052(1995).  
CC -1- FUNCTION: BIFUNCTIONAL ENZYME THAT POSSESSES CHORISMATE SYNTHASE  
CC AND INTRINSIC FLAVIN REDUCTASE ACTIVITY, IT USES NADPH TO REDUCE  
CC FMN.  
CC -1- CATALYTIC ACTIVITY: 5-O-(1-CARBOXYVINYLL)-3-PHOSPHOSHIKIMATE -  
CC CHORISMATE + ORTHOPHOSPHATE.  
CC -1- COFACTOR: REDUCED FLAVIN.  
CC -1- PATHWAY: SEVENTH STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN  
CC THE BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).



CC -!- SIMILARITY: BELONGS TO THE CHORISMATE SYNTHASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U25818; AAC49056.1; -.  
DR InterPro; IPR000453; -.  
DR Pfam; PF01264; Chorismate\_synt; 1.  
DR PROSITE; PS00787; CHORISMATE\_SYNTHASE\_1; 1.  
DR PROSITE; PS00788; CHORISMATE\_SYNTHASE\_2; 1.  
DR PROSITE; PS00789; CHORISMATE\_SYNTHASE\_3; 1.  
DR Lyase; Aromatic amino acid biosynthesis; Oxidoreductase; NADP;  
KW Multifunctional enzyme.  
FT NP\_BIND 260 291 NADPH (POTENTIAL).  
SQ SEQUENCE 432 AA; 45967 MW; 201A0B525C406F0C CRC64;  
  
Query Match 54.8%; Score 34; DB 1; Length 432;  
Best Local Similarity 77.8%; Pred. No. 59;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 EYVPSQHID 10  
DB 274 EYVPSIHND 282  
|||||  
-----  
RESULT 24  
ID SR54\_THEAC STANDARD; PRT; 456 AA.  
AC Q9HKTO;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE SIGNAL RECOGNITION 34 KDA PROTEIN (SRP54).  
GN SRP54 OR TA0515.  
OS Thermoplasma acidophilum.  
OC Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaceae;  
OC Thermoplasma;  
OX NCBI\_TaxID=2303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 1728;  
RX MEDLINE=20479972; PubMed=11029001;  
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;  
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma  
RT acidophilum";  
RT Nature 407:508-513(2000).  
CC -!- FUNCTION: BINDS TO THE SIGNAL SEQUENCE OF PRESECRETORY PROTEIN  
CC WHEN THEY EMERGE FROM THE RIBOSOMES (BY SIMILARITY).  
CC -!- SUBUNIT: ARCHAEL SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA  
CC MOLECULE OF 300 NUCLEOTIDES AND TWO PROTEIN SUBUNITS: SRP54 AND  
CC SRP19 (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
CC -!- DOMAIN: HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN BINDS GTP; THE  
CC M-DOMAIN BINDS THE 7S RNA IN PRESENCE OF SRP19 AND ALSO BINDS THE  
CC SIGNAL SEQUENCE (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AL445064; CAC11655.1; -.  
-----

DR PROSITE; PS00300; SRP54; 1.  
KW Signal recognition particle; GTP-binding; RNA-binding.  
FT DOMAIN 1 289 G-DOMAIN (BY SIMILARITY).  
FT DOMAIN 290 456 M-DOMAIN (BY SIMILARITY).  
FT NP\_BIND 104 111 GTP (BY SIMILARITY).  
FT NP\_BIND 184 188 GTP (BY SIMILARITY).  
FT NP\_BIND 242 245 GTP (BY SIMILARITY).  
SQ SEQUENCE 456 AA; 51080 MW; E5650FA972DA3EA7 CRC64;  
  
Query Match 54.8%; Score 34; DB 1; Length 456;  
Best Local Similarity 75.0%; Pred. No. 62;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 3 VPGSQHID 10  
DB 353 IPGSQKID 360  
|||||  
-----  
RESULT 25  
ID IL7R\_MOUSE STANDARD; PRT; 459 AA.  
AC P16872;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE INTERLEUKIN-7 RECEPTOR ALPHA CHAIN PRECURSOR (IL-7R-ALPHA).  
GN IL7R.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90198875; PubMed=2317865;  
RA Goodwin R.G., Friend D., Ziegler S.F., Jerzy R., Falk B.A.,  
RA Gimpel S., Cosman D., Dover S.K., March C.J., Namen A.E., Park L.S.;  
RT "Cloning of the human and murine interleukin-7 receptors:  
RT demonstration of a soluble form and homology to a new receptor  
RT superfamily";  
RT Cell 60:941-951(1990).  
CC -!- FUNCTION: RECEPTOR FOR INTERLEUKIN-7.  
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA CHAIN AND A COMMON GAMMA CHAIN.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -!- TISSUE SPECIFICITY: SPLEEN; THYMUS; AND FETAL LIVER.  
CC -!- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.  
CC -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M29697; AAA39304.1; -.  
DR PIR; D34791; D34791.  
DR MGD; MGI:96562; IL7r.  
DR InterPro; IPR000950; -.  
DR InterPro; IPR001777; -.  
DR InterPro; IPR002465; -.  
DR Pfam; PF00041; fn3; 1.  
DR PROSITE; PS01355; HEMATOPO\_REC\_S\_F1; 1.  
KW Receptor; Transmembrane; Glycoprotein; Phosphorylation; Signal.  
FT SIGNAL 1 20  
FT CHAIN 21 459 INTERLEUKIN-7 RECEPTOR ALPHA CHAIN.  
FT DOMAIN 21 239 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 240 264 POTENTIAL.  
FT DOMAIN 265 459 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 128 225 FIBRONECTIN TYPE-III.  
FT DOMAIN 184 189 SER/THR-RICH.  
FT CARBOHYD 60 N-LINKED (GLCNAC. . .) (POTENTIAL).  
-----

FT CARBOHYD 115 115 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT MOD.RES 282 282 PHOSPHORYLATION (BY PKC) (POTENTIAL).  
 SQ SEQUENCE 459 AA; 51704 MW; CC06A5CE95543849 CRC64;

Query Match 54.8%; Score 34; DB 1; Length 459;  
 Best Local Similarity 66.7%; Pred. No. 63;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VEVFGSQHI 9  
 :||| ||||:  
 Db 46 LEVDGSQHL 54

Search completed: July 16, 2001, 16:44:03  
 Job time: 462 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 16:43:36 ; Search time 57.41 Seconds  
(without alignments)  
27.655 Million cell updates/sec

Title: US-09-786-648-3

Perfect score: 62

Sequence: 1 VEVPGSQHDSQ 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

SPTREMBL\_16:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organella:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.rodent:\*  
12: sp.unclassified:\*  
13: sp.vertebrate:\*  
14: sp.virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	100.0	103	Q9R646	Q9R646 vibrio chol
2	62	100.0	124	Q57193	Q57193 vibrio chol
3	62	100.0	124	Q56635	Q56635 vibrio chol
4	62	100.0	124	Q9RP15	Q9RP15 vibrio chol
5	43	69.4	395	Q9NKD5	Q9NKD5 drosophila
6	40	64.5	242	Q86582	Q86582 streptomyce
7	40	64.5	254	Q31335	Q31335 bacillus ce
8	39	62.9	192	Q9NP06	Q9NP06 homo sapien
9	39	62.9	201	Q9RJZ8	Q9RJZ8 streptomyce
10	39	62.9	260	Q9MA16	Q9MA16 arabidopsis
11	39	62.9	272	Q9UG04	Q9UG04 homo sapien
12	39	62.9	353	Q9RVJ7	Q9RVJ7 deinococcus
13	39	62.9	412	Q9Y2B3	Q9Y2B3 homo sapien
14	39	62.9	428	Q9LGM2	Q9LGM2 oryza sativ
15	38.5	62.1	565	Q02511	Q02511 vitis vinif
16	37	59.7	331	Q01891	Q01891 drosophila
17	37	59.7	354	Q01437	Q01437 schizosacch
18	37	59.7	367	Q080418	Q080418 nicotiana t
19	37	59.7	508	Q99987	Q99987 homo sapien

20	37	59.7	550	10	Q04636	O04636 arabidopsis
21	36	58.1	325	4	Q9HB36	Q9HB36 homo sapien
22	36	58.1	373	2	Q9XD79	Q9XD79 streptomyce
23	36	58.1	382	5	Q9V7M7	Q9V7M7 drosophila
24	36	58.1	389	5	Q9U0W5	Q9U0W5 leishmania
25	36	58.1	527	14	Q9E226	Q9E226 helicoverpa
26	36	58.1	551	10	Q9SQ79	Q9SQ79 pinus taeda
27	36	58.1	832	14	Q9YFV3	Q9YFV3 hepatitis b
28	36	58.1	832	14	Q9YPU9	Q9YPU9 hepatitis b
29	36	58.1	832	14	Q9DUH5	Q9DUH5 hepatitis b
30	36	58.1	832	14	Q9DUH1	Q9DUH1 hepatitis b
31	36	58.1	843	14	Q91529	Q91529 hepatitis b
32	36	58.1	843	14	Q9QMN7	Q9QMN7 hepatitis b
33	36	58.1	855	4	Q9HCA3	Q9HCA3 homo sapien
34	36	58.1	855	4	Q9H3S0	Q9H3S0 homo sapien
35	35	56.5	99	8	Q37764	Q37764 mytilus edu
36	35	56.5	246	5	Q9V9U6	Q9V9U6 drosophila
37	35	56.5	259	2	Q9K6C4	Q9K6C4 bacillus ha
38	35	56.5	284	2	Q9JW19	Q9JW19 neisseria m
39	35	56.5	363	5	Q76806	Q76806 pisaster oc
40	35	56.5	377	5	Q17446	Q17446 caenorhabdi
41	35	56.5	380	1	Q9HHD0	Q9HHD0 pyrococcus
42	35	56.5	415	5	Q94140	Q94140 caenorhabdi
43	35	56.5	427	13	Q42099	Q42099 cyprinus ca
44	35	56.5	443	5	O01797	O01797 caenorhabdi
45	35	56.5	501	5	Q27038	Q27038 theileria p

#### ALIGNMENTS

#### RESULT 1

Q9R646 ID Q9R646 PRELIMINARY; PRT; 103 AA.  
AC Q9R646;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
RT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
DE CHOLERA-LIKE ENTEROTOXIN B SUBUNIT.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN [1]  
RP SEQUENCE  
RX MEDLINE=95303036; PubMed=7783690;  
RA Nakashima K., Eguchi Y., Nakasone N.;  
RT "Characterization of an enterotoxin produced by Vibrio cholerae  
RT O139.";  
RL Microbiol. Immunol. 39:87-94(1995).  
DR HSP; P01556; LXC.  
DR InterPro; IPR001835;  
DR Pfam; PF01376; Enterotoxin\_B; 1.  
DR PRINTS; PRO0772; ENTEROTOXINB.  
SQ SEQUENCE 103 AA; 11645 MW; 992A05C077F2F70E CRC64;

Query Match 100.0%; Score 62; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.00031;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

#### QY 1 VEVPGSQHDSQ 12

|||||  
Db 50 VEVPGSQHDSQ 61

#### RESULT 2

Q57193 ID Q57193 PRELIMINARY; PRT; 124 AA.  
AC Q57193;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
DE CHOLERA TOXIN B PROTEIN (CTB) PRECURSOR (CTB).

GN CTXB.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CLASSICAL STRAIN 569B;  
 RX MEDLINE=91355224; PubMed=1883840;  
 RA Dams E., De Wolf M., Dierick W.;  
 RT "Nucleotide sequence analysis of the CT operon of the Vibrio cholerae  
 RT classical strain 569B.";  
 RL Biochim. Biophys. Acta 1090:139-141(1991).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-CLASSICAL BIOTYPE 569B;  
 RC SHI C., Cao C., Zhang J., Ma Q.;  
 RL Chin. Biochem. J. 9:395-399(1993).  
 [3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-CLASSICAL BIOTYPE 569B;  
 RA Xu L.;  
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X58785; CAA41591.1; -;  
 DR EMBL; U25679; AAC34728.1; -;  
 DR EMBL; A00931; CAA00098.1; -;  
 DR HSSP; P01556; 2CHB.  
 DR InterPro; IPR001835; -;  
 DR Pfam; PF01376; Enterotoxin\_B; 1.  
 DR PRINTS; PR00772; ENTEROTOXINB.  
 DR ProDom; PD012805; -; 1.  
 KW Signal.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 124 CHOLERA TOXIN B PROTEIN (CTB).  
 FT SEQUENCE 124 AA; 13919 MW; D6BF83FF7924EA3 CRC64;  
 SQ SEQUENCE 124 AA; 13919 MW; D6BF83FF7924EA3 CRC64;  
 Query Match 100.0%; Score 62; DB 2; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.00037;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VEVPGSQHDSQ 12  
 Db 71 VEVPGSQHDSQ 82  
 RESULT 3  
 Q56635 PRELIMINARY; PRT; 124 AA.  
 ID Q56635;  
 AC Q56635;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE CHOLERA TOXIN PRECURSOR.  
 GN CTXB.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-S7;  
 RA Yamamoto K., Do V.G., Xu M., Iida T., Miwatani T., Albert M.J.,  
 RT Honda T.;  
 RT "Comparison of cholera toxin genes (ctxAB) of non-O1 vibrio cholerae  
 RT strains 854 (O139-Bengal) and S7 (O37) from two outbreaks.";  
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; D30052; BAA06289.1; -;  
 DR HSSP; P01556; 2CHB.  
 DR InterPro; IPR001835; -;  
 DR Pfam; PF01376; Enterotoxin\_B; 1.  
 DR PRINTS; PR00772; ENTEROTOXINB.  
 DR ProDom; PD012805; -; 1.  
 KW Signal.  
 FT SIGNAL 1 21 POTENTIAL.

SQ SEQUENCE 124 AA; 13871 MW; 3F87B2F297953179 CRC64;  
 Query Match 100.0%; Score 62; DB 2; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.00037;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VEVPGSQHDSQ 12  
 Db 71 VEVPGSQHDSQ 82  
 RESULT 4  
 Q9RP15 PRELIMINARY; PRT; 124 AA.  
 ID Q9RP15;  
 AC Q9RP15;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE CHOLERA ENTEROTOXIN B-SUBUNIT.  
 GN CTXB.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-KNIH002;  
 RA Shin H.J., Park Y.C., Kim Y.C.;  
 RT "Cloning and nucleotide sequence analysis of the virulence gene  
 RT cassette from Vibrio cholerae KNIH002 isolated in Korea.";  
 RL Misainmurah Hoiji 35:205-210(1999).  
 DR EMBL; AF175708; AAD51360.1; -;  
 DR HSSP; P01556; 2CHB.  
 DR InterPro; IPR001835; -;  
 DR Pfam; PF01376; Enterotoxin\_B; 1.  
 DR PRINTS; PR00772; ENTEROTOXINB.  
 DR SEQUENCE 124 AA; 13905 MW; 23BF83FF793E5B9 CRC64;  
 SQ SEQUENCE 124 AA; 13905 MW; 23BF83FF793E5B9 CRC64;  
 Query Match 100.0%; Score 62; DB 2; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.00037;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VEVPGSQHDSQ 12  
 Db 71 VEVPGSQHDSQ 82  
 RESULT 5  
 Q9NKD5 PRELIMINARY; PRT; 395 AA.  
 ID Q9NKD5;  
 AC Q9NKD5;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE HYPOTHETICAL 43.6 KDA PROTEIN.  
 GN BG:DS01514.3.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-Y, CN BW SP;  
 RX MEDLINE=99403001; PubMed=10471707;  
 RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,  
 RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,  
 RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,  
 RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,  
 RA Ceiniker S., Rubin G.W.;  
 RT "An exploration of the sequence of a 2.9-Mb region of the genome of  
 RT Drosophila melanogaster: the Adh region.";

```

RN  Genetics 153:179-219(1999).
RP  SEQUENCE FROM N.A.
RC  STRAIN=Y, CN BW SP;
RA  Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA  Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA  Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA  Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA  Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA  Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA  Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,
RA  Zieran L.L., Rubin G.M.;
RL  Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AF003408; AAF44847.1; -
DR  FlyBase; FBgn0028907; BG:DS01514.3.
KW  Hypothetical protein.
SQ  SEQUENCE 395 AA; 43561 MW; AE4F1CC4ADD3DA73 CRC64;

Query Match 69.4%; Score 43; DB 5; Length 395;
Best Local Similarity 72.7%; Pred. No. 4.7;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY  1 VEVPGSQHIDS 11
DB  330 VSVPGSTHIDA 340
      | |||| |||:
      | |||| |||:

RESULT 6
ID  O86582 PRELIMINARY; PRT; 242 AA.
AC  O86582;
DT  01-NOV-1998 (TREMBlrel. 08, Created)
DT  01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DE  01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE  HYPOTHETICAL 26.0 KDA PROTEIN.
GN  SC2A11.21C.
OS  Streptomyces coelicolor.
OC  Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC  Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX  NCBI_TaxID=1902;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=A3(2);
RC  Murphy L., Harris D.;
RA  Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=A3(2);
RA  Parkhill J., Bartell B.G., Rajandream M.A.;
RL  Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN=A3(2);
RX  MEDLINE=97000351; PubMed=8843436;
RA  Redenbach M., Kiser H.M., Denapate D., Eichner A., Cullum J.,
RA  Kinashi H., Hopwood D.A.;
RT  "A set of ordered cosmids and a detailed genetic and physical map for
RT  the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL  Mol. Microbiol. 21:77-96(1996).
DR  EMBL; AL031184; CAA20190.1; -
DR  InterPro; IPR002502; -
DR  Pfam; PF01510; Amidase_2; 1.
KW  Hypothetical protein.
SQ  SEQUENCE 242 AA; 26018 MW; 6DD10FF18A2EC544 CRC64;

Query Match 64.5%; Score 40; DB 2; Length 242;
Best Local Similarity 70.0%; Pred. No. 10;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY  1 VEVPGSQHID 10
      | ||||: | |

```

```

Db  204 VEVPGTDHTD 213

RESULT 7
Q31335 PRELIMINARY; PRT; 254 AA.
AC  Q31335;
DT  01-JAN-1998 (TREMBlrel. 05, Created)
DT  01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT  01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE  ATP-DEPENDENT NUCLEASE SUBUNIT A (FRAGMENT).
GN  ADDA.
OS  Bacillus cereus.
OC  Bacteria; Firmicutes; Bacillus/Clostridium group;
OC  Bacillus/Staphylococcus group; Bacillus.
OX  NCBI_TaxID=1396;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=ATCC 10987;
RA  Kolsto A.B., Okstad O.A., Lindback T., Hegna I., Lagreid A.,
RA  Rishovd A.L.;
RL  Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR  EMBL; Y11217; CAA72103.1; -
DR  InterPro; IPR00212; -
DR  Pfam; PF00580; UvrD-helicase; 1.
FT  NON_TER 1 254
FT  NON_TER 1 254
SQ  SEQUENCE 254 AA; 29212 MW; 24392E11338DD99EA CRC64;

Query Match 64.5%; Score 40; DB 2; Length 254;
Best Local Similarity 58.3%; Pred. No. 11;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY  1 VEVPGSQHIDS 12
DB  35 IDEPGSHIRKQ 46
      :: ||||| |
      :: ||||| |

RESULT 8
Q3NPQ6 PRELIMINARY; PRT; 192 AA.
AC  Q3NPQ6;
DT  01-OCT-2000 (TREMBlrel. 15, Created)
DT  01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT  01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE  LLPL, LCAT-LIKE LYSOPHOSPHOLIPASE (FRAGMENT).
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Pluvinet R., Estivill X., Escarceller M., Sumoy L.;
RL  Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RA  Aufray C., Ansorge W., Ballabio A., Estivill X., Gibson K.,
RA  Lehrach H., Pouska A., Lundeberg J.;
RT  "The European IMAGE consortium for integrated Molecular analysis of
RT  human gene transcripts.";
RL  Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AL389957; CAB97531.1; -
FT  NON_TER 1 192
SQ  SEQUENCE 192 AA; 21609 MW; 04A7AB8CB344F213 CRC64;

Query Match 62.9%; Score 39; DB 4; Length 192;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY  2 EYVPGSHID 10
      | : |||: | |

```

Db 166 ELPGEHIE 174

## RESULT 9

Q9RJZ8 PRELIMINARY; PRT; 201 AA.  
AC Q9RJZ8;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
DE PUTATIVE AMIDASE.  
GN SCF37.03.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
[1]  
RN Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Oliver K., Harris D.;  
RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Redenbach M., Klesner H.M., Denaparte D., Eichner A., Cullum J.,  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Klesner H.M., Denaparte D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmid and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
RL Mol. Microbiol. 21:77-96(1996).  
DR EMBL; AL133210; CAB61584.1; -;  
DR InterPro; IPR002502; -;  
DR Pfam; PF01510; Amidase\_2; 1.  
SQ SEQUENCE 201 AA; 22749 MW; B8EF477E06A20468 CRC64;

Query Match 62.9%; Score 39; DB 2; Length 201;

Best Local Similarity 77.8%; Pred. No. 13;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVPGSOHID 10

|:|||||:

Db 176 EVPGSDHTD 184

## RESULT 10

Q9MA16 PRELIMINARY; PRT; 260 AA.  
AC Q9MA16;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE F20B17.2  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
[1]  
RN Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.  
RP SEQUENCE FROM N.A.  
RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,  
RA Shinn P., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,  
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,  
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,  
RA Ecker J.R.;  
RT "Genomic sequence for Arabidopsis thaliana BAC F20B17 from chromosome

I.,"  
RT Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.  
RL [2]

RN SEQUENCE FROM N.A.

RP Ecker J.R.;

RA Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.  
RN [3]

RP SEQUENCE FROM N.A.

RA Ecker J.R.;

RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.  
RN [4]

RP SEQUENCE FROM N.A.

RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
RA Khan S., Kim C., Altafi H., Bei Q., Chin C., Chiou J., Choi E.,  
RA Conn L., Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,  
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,  
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
RA Theologis A., Ecker J.R.;  
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AC010793; AAF68106.1; -;  
SQ SEQUENCE 260 AA; 29128 MW; B149F22073AA0B92 CRC64;

Query Match 62.9%; Score 39; DB 10; Length 260;

Best Local Similarity 50.0%; Pred. No. 17;

Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VEVPQSOHDSQ 12

|:|||||:

Db 65 KVPGKQHVSEK 76

## RESULT 11

Q9UG04 PRELIMINARY; PRT; 272 AA.  
AC Q9UG04;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE HYPOPHETICAL 31.0 KDA PROTEIN.  
GN DKFZP564A0122.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RN Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RA Wambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;  
RL Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AL110209; CAB53675.1; -;  
DR InterPro; IPR003386; -;  
DR Pfam; PF02450; LACT; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 272 AA; 31016 MW; ACCC5E1680D7A720 CRC64;

Query Match 62.9%; Score 39; DB 4; Length 272;

Best Local Similarity 66.7%; Pred. No. 18;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 EVPGSOHID 10

|:|||||:

Db 246 ELPGEHIE 254

## RESULT 12

Q9RVJ7 PRELIMINARY; PRT; 353 AA.

AC Q9RVJ7;

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

```

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE (S)-2-HYDROXY-ACID OXIDASE.
GN DR1031.
OC Deinococcus radiodurans.
OS Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
ON NCBI_TaxID=1299;
RX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=RL;
RC MEDLINE=20036896; PubMed=10567266;
RX White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AF001954; AAF10604.1; -.
DR HSSP; P05414; IGOX.
DR TIGR; DR1031; -.
DR InterPro; IPR000262; -.
DR InterPro; IPR003009; -.
DR Pfam; PF01070; FMN_dh; 1.
DR PROSITE; PS00557; FMN_HYDROXY-ACID_DH; 1.
SQ SEQUENCE 353 AA; 37877 MW; 14FB78FAE2E18C8D CRC64;

Query Match 62.9%; Score 39; DB 2; Length 353;
Best Local Similarity 75.0%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 VPGSQHID 10
   |||||:|
Db 189 VPGSEHLD 196

RESULT 13
QY2B3 ID QY2B3 PRELIMINARY; PRT; 412 AA.
AC QY2B3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE LCAT-LIKE PROTEIN (LLPL).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99194552; PubMed=10092508;
RA Taniyama Y., Shibata S., Kita S., Horikoshi K., Shirafuji H.,
RA Sumino Y., Fujino M.;
RT "Cloning and expression of a novel lysophospholipase which
RT structurally resembles lecithin cholesterol acyltransferase.";
RL Biochem. Biophys. Res. Commun. 257:50-56(1999).
DR EMBL; AB017494; BAA76877.1; -.
DR InterPro; IPR003386; -.
DR Pfam; PF02450; LACT; 1.
SQ SEQUENCE 412 AA; 46657 MW; 1FEA8A5783AF050A CRC64;

Query Match 62.9%; Score 39; DB 4; Length 412;
Best Local Similarity 66.7%; Pred. No. 28;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 EVPGSQHID 10
   |||||:|
Db 386 ELPGSEHIE 394

```

```

RESULT 14
Q9LGM2 ID Q9LGM2 PRELIMINARY; PRT; 428 AA.
AC Q9LGM2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ESTS AU056822(S20908).
OS Oryza sativa (rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
OX Oryza.
ON NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone: P0041E11.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone: P0433F09.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002521; BAA96762.1; -.
DR EMBL; AP002539; BAB08201.1; -.
DR InterPro; IPR001552; -.
DR Pfam; PF00441; Acyl-CoA_dh; 1.
DR PROSITE; PS00073; ACYL_COA_DH_2; UNKNOWN1.
SQ SEQUENCE 428 AA; 46132 MW; 8D34E3698AE6367 CRC64;

Query Match 62.9%; Score 39; DB 10; Length 428;
Best Local Similarity 53.6%; Pred. No. 29;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVPGSQHIDSQ 12
   |||||:|
Db 187 KVPGGHIDGQ 197

RESULT 15
Q22511 ID Q22511 PRELIMINARY; PRT; 565 AA.
AC Q22511;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE GLUTATHIONE REDUCTASE (NADPH) (EC 1.6.4.2) (FRAGMENT).
GN GOR.
OS Vitis vinifera (Grape).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Vitaceae; Vitis.
OX NCBI_TaxID=29760;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. THOMPSON SEEDLESS (CLONE 2A); TISSUE=FRUIT;
RA Cassol T., Adams D.O.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -| COFACTOR: FAD (BY SIMILARITY).
CC -| SIMILARITY: TO PYRIDINE NUCLEOTIDE-DISULPHIDE OXIDOREDUCTASES
CC CLASS-I.
DR EMBL; AF019907; AAB70837.1; -.
DR HSSP; P00390; IALG.
DR Mendel; 26381; Vitvi; 1190; 26381.
DR InterPro; IPR000103; -.
DR InterPro; IPR001100; -.
DR InterPro; IPR001327; -.
DR Pfam; PF00070; pyr_redox; 1.

```

DR PRINTS; PRO0368; FADPFR.

DR PRINTS; PRO0411; PNDRTASEI.

DR PRINTS; PRO0469; PNDRTASEII.

DR PROSITE; PS00076; PYRIDINE\_REDOX\_1; 1.

FW FAD; Flavoprotein; Oxidoreductase; Redox-active center.

KT NON\_TER 1

SQ SEQUENCE 565 AA; 60695 MW; B26113AE09A121DE CRC64;

Query Match 62.1%; Score 38.5; DB 10; Length 565;

Best Local Similarity 72.7%; Pred. NO. 48;

Matches 8; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

QY 2 EYVPSGH-IDS 11

Db 239 EIPGSEHAIDS 249

RESULT 16

O18391

ID O18391 PRELIMINARY; PRT; 331 AA.

AC O18391;

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE PROBABLE SERINE HYDROLASE (EC 3.1.-.-) (KRAKEN PROTEIN).

GN KRAKEN OR CG3943.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

[1]

RN SEQUENCE FROM N.A.

RP TISSUE=EMBRYO;

RC MEDLINE=20196006; PubMed=9831651;

RX Chan H.Y.E., Harris S.J., O'Kane C.J.;

RT "Identification and Characterization of kraken, a gene encoding a

RT putative hydrolytic enzyme in Drosophila melanogaster.";

RL Gene 222:195-201(1998).

[2]

RN SEQUENCE FROM N.A.

RP STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-pfannkuch C., Baldwin D.,

RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cusum S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.C.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svrls R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

CC -1- FUNCTION: MAY HAVE A ROLE IN DETOXIFICATION AND DIGESTION DURING

CC EMBRYOGENESIS AND LARVAL DEVELOPMENT.

CC -1- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED BEFORE EMBRYONIC STAGE

CC 11. AT STAGE 11, EXPRESSION IS CONCENTRATED IN THE FOREGUT AND

CC POSTERIOR MIDGUT. BY STAGE 15, IN GASTRIC CAECAE, PHARYNX,

CC POSTERIOR SPIRACLES AND ANTERIOR EDGE OF MIDGUT. AT THE END OF

CC EMBRYOGENESIS, EXPRESSION IS CONFINED TO GASTRIC CAECAE. DURING

CC THIRD INSTAR LARVAE, EXPRESSED AT LOW LEVELS IN GASTRIC CAECAE,

CC MIDGUT AND HINDGUT AND HIGH LEVEL IN FAT BODY.

CC -1- DEVELOPMENTAL STAGE: PROBABLY EXPRESSED BOTH MATERNALLY AND

CC ZYGOTICALLY.

CC -1- SIMILARITY: WEAK. TO FAMILY OF ESTERASES THAT GROUPS TOGETHER

CC PSEUDOMONA TROPINESTERASE, DMPD; TODF AND XYL.

DR EMBL; AJ000516; CA04153.1; -.

DR EMBL; AE003588; AAF51445.1; -.

DR FlyBase; FBgn0020545; kraken.

DR InterPro; IPR000073; -.

DR InterPro; IPR000379; -.

DR InterPro; IPR000734; -.

DR Pfam; PF00561; anhydrolase; 1.

DR PROSITE; PS00120; LIPASE\_SER; 1.

KW Detoxification; Developmental protein; Digestion; Hydrolase;

KW Serine esterase.

FT ACT\_SITE 138 138 BY SIMILARITY.

SQ SEQUENCE 331 AA; 37093 MW; 1FE39BC42AED4E69 CRC64;

Query Match 59.7%; Score 37; DB 5; Length 331;

Best Local Similarity 66.7%; Pred. No. 52;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VEVPGSQHI 9

Db 304 VEVPGTHHL 312

RESULT 17

O14347

ID O14347 PRELIMINARY; PRT; 354 AA.

AC O14347;

DT 01-JUN-1998 (TREMBLrel. 06, Created)

DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE PUTATIVE ZUOTIN-LIKE PROTEIN C30D10.01 (FRAGMENT).

GN SPBC30D10.01.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC NCBI\_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RA Wood V., Rajandream M.A., Barrell B.G., Lauber J., Hilbert H.,

RA Duesterhoeft A.;

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: 2-DNA BINDING PROTEIN. COULD BE INVOLVED IN CHROMOSOME

CC ORGANIZATION (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -1- SIMILARITY: CONTAINS A DNAJ-LIKE DOMAIN.

DR EMBL; Z97992; CAB10796.1; -.

DR HSSP; P25685; 1HDJ.

DR InterPro; IPR001623; -.

DR Pfam; PF00226; DnaJ; 1.



DR PROSITE; PS00636; DNAL1; 1.  
 DR PROSITE; PS00076; DNAL2; 1.  
 DR SMART; SM00271; DnaJ; 1.  
 KW Hypothetical protein; Chaperone; DNA-binding; Nuclear protein.  
 FT NON\_TER 1 1 DNAL-LIKE.  
 FT DOMAIN 10 81 ALA/LYS-RICH.  
 FT DOMAIN 217 267  
 SQ SEQUENCE 354 AA; 40290 MW; 6071B58A3B60F558 CRC64;  
 Query Match 59.7%; Score 37; DB 3; Length 354;  
 Best Local Similarity 50.0%; Pred. No. 56;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 EVPGSQHDS 11  
 Db 283 DVPSAEHVD 292  
 RESULT 18  
 ID O80418 PRELIMINARY; PRT; 367 AA.  
 AC O80418;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE NTC16 PROTEIN.  
 GN NTC16.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae I;  
 OC Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tanaka-Ueguchi M., Itoh H., Oyama N., Koshioka M., Matsuoka M.;  
 RT "Over-expression of a tobacco homeobox gene, NTH15, decreases the  
 RT expression of a gibberellin biosynthetic gene encoding GA 20-  
 RT oxidase.";  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB015084; BAA31690.1;  
 DR Mendel; 31081; Nicta; 2972; 31081.  
 DR InterPro; IPR002419;  
 DR Pfam; PF00671; Fe\_Asc-oxidore; 1.  
 SQ SEQUENCE 367 AA; 42170 MW; 923BC90B3BBAC05 CRC64;  
 Query Match 59.7%; Score 37; DB 10; Length 367;  
 Best Local Similarity 87.5%; Pred. No. 58;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 EVPGSQHI 9  
 Db 149 EVPSSQHI 156  
 RESULT 19  
 Q99987  
 ID Q99987 PRELIMINARY; PRT; 508 AA.  
 AC Q99987;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE VRK2, COMPLETE CDS.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RX MEDLINE=98008921; PubMed=9344656;  
 RA Nezu J., Oku A., Jones M.H., Shimane M.;

RT "Identification of two novel human putative serine/threonine kinases,  
 RT VRK1 and VRK2, with structural similarity to vaccinia virus B1R  
 RT kinase.";  
 RL Genomics 45:327-331(1997).  
 CC 1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; AB000450; BAA19109.1;  
 DR HSP; Q06486; ICKI.  
 DR InterPro; IPR000719;  
 DR InterPro; IPR002290;  
 DR Pfam; PF00069; pkinase; 2.  
 DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR SMART; SM00220; S\_TKc; 1.  
 KW ATP-binding; Serine/threonine-protein kinase; Transferase.  
 SQ SEQUENCE 508 AA; 58126 MW; 157FBF8F48511AF4 CRC64;  
 Query Match 59.7%; Score 37; DB 4; Length 508;  
 Best Local Similarity 58.3%; Pred. No. 82;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 VEVPGSQHDSQ 12  
 Db 334 VHTPNSQKVDSQ 345  
 RESULT 20  
 O04636 PRELIMINARY; PRT; 550 AA.  
 ID O04636;  
 AC O04636;  
 DT 01-JUL-1997 (TREMBLrel. 04, Created)  
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE SIMILARITY TO GATA-TYPE ZINC FINGERS.  
 GN A\_I002P16.9.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Miller N., Beck C., Kramer J.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Waterston R.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF007270; AAB61058.1;  
 DR Mendel; 17174; Arath; 2663; 17174.  
 DR InterPro; IPR000679;  
 DR Pfam; PF00320; GATA; 1.  
 DR PROSITE; PS01114; GATA\_2N\_FINGER\_2; 2.  
 DR SMART; SM00401; ZnF\_GATA; 1.  
 SQ SEQUENCE 550 AA; 60856 MW; 488A05F20846091D CRC64;  
 Query Match 59.7%; Score 37; DB 10; Length 550;  
 Best Local Similarity 66.7%; Pred. No. 90;  
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 VEVPGSQHDSQ 12  
 Db 260 VTCSGSQHIDFQ 271  
 RESULT 21  
 Q9HB36 PRELIMINARY; PRT; 325 AA.  
 ID Q9HB36;  
 AC Q9HB36;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
 DE SERINE PROTEASE SMC19 (FRAGMENT)  
 GN ST14.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cao J., Fan W., Zheng S.;  
 RT "Genomic analysis of a novel human serine protease SMC19."  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF283256; AAG13949.1; -;  
 KW Protease.  
 FT NON\_TER 325 325  
 FT NON\_TER 325 325  
 SQ SEQUENCE 325 AA; 35896 MW; F6A7468C1B26B64F CRC64;  
  
 Query Match 58.1%; Score 36; DB 4; Length 325;  
 Best Local Similarity 55.6%; Pred. No. 79;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
 QY 1 VEVPGSQHI 9  
 Db :||| :|||  
 Db 31 IEVPPNQHV 39  
  
 RESULT 22  
 Q9XD79 PRELIMINARY; PRT; 373 AA.  
 AC Q9XD79;  
 DT 01-NOV-1999 (TReMBLrel. 12, Created)  
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
 DE 4-CARBOXYMUCONOLACTONE DECARBOXYLASE/3-OXOADIPATE ENOL-LACTONE  
 DE HYDROLASE.  
 GN PCAL.  
 OS Streptomyces sp. 2065.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=86383;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Yang K., Iwagami S., Davies J.E.;  
 RT "A protocatechuate catabolic gene cluster cloned from Streptomyces sp.  
 RT 2065."  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.  
 DR EMBL; AF109386; AAD40815.1; -;  
 DR InterPro; IPR000073; -;  
 DR InterPro; IPR000379; -;  
 DR InterPro; IPR003089; -;  
 DR Pfam; PF00561; abhydrolase; 1.  
 DR PRINTS; PR00111; ABHYDROLASE.  
 KW Hydrolase.  
 SQ SEQUENCE 373 AA; 39583 MW; F11D3017D7A524DC CRC64;  
  
 Query Match 58.1%; Score 36; DB 2; Length 373;  
 Best Local Similarity 55.6%; Pred. No. 92;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
 QY 1 VEVPGSQHI 9  
 Db :||| :|||  
 Db 225 VEIPGASHL 233  
  
 RESULT 23  
 Q9V7M7 PRELIMINARY; PRT; 382 AA.  
 ID Q9V7M7

AC Q9V7M7;  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
 DE CG7750 PROTEIN.  
 GN CG7750.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydriidae; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brotter P.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieri S., Fleischmann W.,  
 RA Fostier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasako P., Lei Y., Levitsky A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J., Wu D.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 DR EMBL; AF003807; AAF38021.1; -;  
 DR FlyBase; FBgn0034107; CG7750.  
 SQ SEQUENCE 382 AA; 43802 MW; DAAC4084BB5E3A98 CRC64;

Query Match 58.1%; Score 36; DB 5; Length 382;  
 Best Local Similarity 54.5%; Pred. No. 94;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 EYVPGSHDSQ 12  
 Db :||| :|||  
 Db 348 ELPGQHKETQ 358

RESULT 24  
 Q9U0W5 PRELIMINARY; PRT; 389 AA.  
 AC Q9U0W5;  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE HYPOTHETICAL 42.7 KDA PROTEIN.  
GN L7276.04.  
OS Leishmania major.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
OX NCBI\_TaxID=5664;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FRIEDLIN;  
RA Zimmermann W., Wambutt R., Ivens A.C., Murphy L., Quail M.,  
RA Rajandream M.A., Barrell B.G.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FRIEDLIN;  
RX MEDLINE=98146435; PubMed=9477341;  
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,  
RA Smith D.F.;  
RT "A physical map of the Leishmania major Friedlin genome."  
RL Genome Res. 8:135-145(1998).  
DR EMBL; AL133436; CAB62821.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 389 AA; 42733 MW; 4C1A0CF31D0DC670 CRC64;

Query Match 58.1%; Score 36; DB 5; Length 389;  
Best Local Similarity 66.7%; Pred. No. 96;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 PGSOHDSQ 12  
DB 107 PGARHIDVQ 115

RESULT 25  
Q9E226 PRELIMINARY; PRT; 527 AA.  
AC Q9E226;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE ORF119-LIKE PROTEIN.  
OS Helicoverpa zea nuclear polyhedrosis virus.  
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
OC Nucleopolyhedrovirus.  
OX NCBI\_TaxID=10468;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Robertson A.P.S.;  
RT "Genetic organization of Helicoverpa zea nuclear polyhedrosis virus in  
RT the region of EcoRI 0,D,L,A and Q.";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF275264; AAG17375.1; -.  
SQ SEQUENCE 527 AA; 59935 MW; E1DF20478A77574C CRC64;

Query Match 58.1%; Score 36; DB 14; Length 527;  
Best Local Similarity 54.5%; Pred. No. 1.3e+02;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VEVPGSQHIDS 11  
DB 47 IEIPGEINIDS 57

Search completed: July 16, 2001, 16:43:37  
Job time: 471 sec

*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 16:35:41 ; Search time 56.06 Seconds  
(without alignments)  
22.710 Million cell updates/sec

Title: US-09-786-648-4

Perfect score: 106

Sequence: 1 GATFQVEVPGSQHDSQKAI 21

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_0601.\*

- 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*
- 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	106	100.0	21	AA1987462	Cholera toxin B/en
2	106	100.0	103	AA1987462	Synthetic cholera
3	106	100.0	103	AA1987462	Heat labile entero
4	106	100.0	103	AA1987462	Cholera toxin B su
5	106	100.0	103	AA1987462	Cholera toxin B su
6	106	100.0	103	AA1987462	Amino acid sequenc
7	106	100.0	118	AA1987462	Cholera Toxin B-su
8	106	100.0	124	AA1987462	B subunit of the h
9	106	100.0	124	AA1987462	Cholera toxin B su
10	106	100.0	124	AA1987462	Amino acid sequenc
11	106	100.0	124	AA1987462	Plant-optimized E.

12	106	100.0	131	AA1987462	Sequence of LT-B-M
13	106	100.0	138	AA1987462	LT-B-CPP fusion pro
14	106	100.0	170	AA1987462	HSV-1 antigen/heat
15	106	100.0	405	AA1987462	Sequence of sub-un
16	103	97.2	103	AA1987462	E. coli heat labil
17	101	95.3	21	AA1987462	ADP-ribosylating t
18	101	95.3	93	AA1987462	Escherichia coli v
19	101	95.3	93	AA1987462	E. coli heat-labli
20	101	95.3	93	AA1987462	Heat labile toxin
21	101	95.3	93	AA1987462	E coli verotoxin-1
22	101	95.3	93	AA1987462	Recombinant exotox
23	101	95.3	134	AA1987462	Recombinant exotox
24	101	95.3	142	AA1987462	Recombinant exotox
25	101	95.3	155	AA1987462	Recombinant exotox
26	101	95.3	163	AA1987462	Recombinant exotox
27	101	95.3	371	AA1987462	Labile toxin (LT-B
28	101	95.3	371	AA1987462	C. jejuni flagelli
29	100	94.3	124	AA1987462	Plant-optimized V.
30	100	94.3	126	AA1987462	GtFB-1/CTB chimeri
31	94	88.7	124	AA1987462	B subunit of CT.
32	94	88.7	461	AA1987462	Adhesin/V.cholerae
33	94	88.7	749	AA1987462	Helicobacter pylor
34	94	88.7	1338	AA1987462	Helicobacter pylor
35	92.5	87.3	47	AA1987462	Sequence of amino
36	89	84.0	46	AA1987462	Network polymer wh
37	88	83.0	41	AA1987462	Network polymer wh
38	87	82.1	461	AA1987462	Adhesin/CTX2A chi
39	80	75.5	26	AA1987462	Sequence of amino.
40	76	71.7	15	AA1987462	CTP3 epitope of th
41	76	71.7	15	AA1987462	Cholera toxin B an
42	76	71.7	23	AA1987462	Residues 50-64 of
43	62	58.5	12	AA1987462	Cholera toxin B/en
44	61	57.5	15	AA1987462	Cholera toxin B su
45	61	57.5	15	AA1987462	Bovine rotavirus V

## ALIGNMENTS

RESULT 1

AA1987462

ID AA1987462 standard; peptide; 21 AA.

XX

AC AA1987462;

XX 03-JUL-2000 (first entry)

DT

XX Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:4.

XX Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;  
XX beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;  
XX adjuvant; immune disorder; diarrhoea.

XX Vibrio cholerae.

XX Escherichia coli.

XX WO200014114-A1.

XX 16-MAR-2000.

XX 07-SEP-1999; 99WO-GB02970.

XX 07-SEP-1998; 98GB-0019484.

XX (UYBR-) UNIV BRISTOL.

XX Williams NA, Hirst TR;

XX WPI; 2000-256943/22.

XX Derivatives of Escherichia coli heat labile enterotoxins useful as  
XX immunomodulators and for treating diarrhea and which do not bind the  
XX glycolipid receptor GM-1.

XX PS Disclosure; Page 15; 62pp; English.

XX CC The invention relates to peptide fragments of the *Escherichia coli* heat labile enterotoxin (Etx) and its closely related homologues, cholera toxin (Ctx) from *Vibrio cholerae* which do not bind to the ubiquitous GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are composed of one A subunit and five identical B subunits. The A subunit is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-riboyltransferase activity, while the B subunits (EtxB and CtxB) facilitate the entry of subunit A into the host cell via the binding and cross-linking of GM-1 receptors. Although GM-1 binding is responsible for some of the effects of Etx and Ctx, it has been found that certain effects of the toxins, such as immunomodulation, are not mediated through GM-1 binding. The peptides of the invention are fragments of the beta-4' alpha-2 loop of EtxB and/or CtxB, exerting the same effects as normal EtxB and CtxB subunits, except that they do not bind or cross link GM-1. They may be used in medicine as an immunomodulator or adjuvant. They may also be used as an inhibitor for toxin-induced diarrhoea. Therefore, the peptides may be used in the production of a composition for treating, preventing and/or modulating a disease associated with an immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-187463 represent preferred peptides of the invention, AAY87460 being particularly preferred.

XX SQ Sequence 21 AA;

Query Match 100.0%; Score 106; DB 21; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2.7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFQVEVPGSQHDSQKKAI 21  
Db 1 GATFQVEVPGSQHDSQKKAI 21

RESULT 2

XX ID AAW04857 standard; Protein; 103 AA.

XX AC AAW04857;

XX DT 21-FEB-1997 (first entry)

XX DE Synthetic cholera toxin B subunit.

XX KW Bordetella pertussis; whooping cough; recombinant construct; cholera toxin B subunit; enzyme; antigen; immunogen; allergen; enzyme inhibitor; hormone; lymphokine; immunoglobulin; toxin; structural protein; receptor; heterologous gene; leader; promoter.

XX OS Synthetic.

XX PN WO9626282-A1.

XX PD 29-AUG-1996.

XX PF 23-FEB-1996; 96WO-CA00107.

XX PR 23-FEB-1995; 95US-0393334.

XX PA (CONN-) CONNAUGHT LAB LTD.

XX PI Klein MH, Loosmore SM, Yacoub RK, Zealey GR;

XX DR WPI; 1996-425088/42.

XX DR N-PSDB; AAT38038.

XX PT Recombinant constructs for expressing and opt. secreting proteins in Bordetella - comprise Bordetella promoter coupled to non-Bordetella, esp. cholera B toxin, gene or coupled to non-Bordetella leader and gene of interest

XX PS Example 1; Figure 1; 61pp; English.

XX CC Recombinant constructs comprising a promoter functional in Bordetella operatively linked to a heterologous gene or a non-Bordetella leader sequence for secretion of a gene product which may or may not be of Bordetella origin, can be used for the expression in Bordetella of enzymes, antigens, immunogens, allergens, enzyme inhibitors, hormones, lymphokines, immunoglobulins or their fragments, toxins, mammalian proteins, structural proteins or receptors. The Bordetella strains are particularly engineered to express the cholera toxin B subunit (this sequence). The promoters used in the constructs are selected from the Bordetella pertussis tox, fha promoters or the high molecular weight (hmw) outer membrane promoter of non typable *Haemophilus influenzae*; leaders used in the constructs are selected from the cholera toxin B leader (CTB-L), the pertussis toxin subunit S1 leader (ST-L) and the pertussin pertactin leader (PRN-L); and genes used in the constructs are selected from a novel synthetic cholera toxin B gene (ctb) and the hmw1 and hmw2 genes of *Haemophilus influenzae*.

XX SQ Sequence 103 AA;

Query Match 100.0%; Score 106; DB 17; Length 103;  
Best Local Similarity 100.0%; Pred. No. 1.7e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFQVEVPGSQHDSQKKAI 21  
Db 45 GATFQVEVPGSQHDSQKKAI 65

RESULT 3

XX ID AAR94939 standard; Protein; 103 AA.

XX AC AAR94939;

XX DT 31-OCT-1996 (first entry)

XX DE Heat labile enterotoxin B subunit (LT-B) *E. coli*.

XX KW Toxin; subunit; vaccine; transgenic plant; immunogen; antigen; adjuvant; immunisation.

XX OS *Escherichia coli*.

XX PN WO9612801-A1.

XX PD 02-MAY-1996.

XX PF 24-OCT-1995; 95WO-US13376.

XX PR 24-OCT-1994; 94US-0328716.

XX PA (TULA) TULANE EDUCATIONAL FUND.

XX PA (TEXA) UNIV TEXAS A & M SYSTEM.

XX PI Arntzen CJ, Clements JD, Haq TA, Mason HS;

XX DR WPI; 1996-230602/23.

XX DR N-PSDB; AAT18799, AAT18800.

XX PT Transgenic plants contg. *E. coli* heat labile enterotoxin subunits used as oral vaccines for animals which consume the plant

XX PS Disclosure; Page 100-101; 130pp; English.

XX CC A transgenic plant comprising or expressing a DNA sequence encoding an immunogenic agent can be used as an oral vaccine for animals. The vaccine is administered by the oral consumption of the plant and provides the first known functional method for immunising animals

CC using transgenic plants, where the plants express bacterial antigens  
 CC that act as both immunogens and adjuvants. The method provides an  
 CC inexpensive production and delivery system for such antigens to  
 CC animals. This is the LT-B Escherichia coli toxin subunit and its  
 CC coding sequence was used in the construction of such a transgenic  
 CC plant. The immunogenic agent preferably comprises the LT-B or CT-B  
 CC (cholera toxin B subunit) or optionally LT-A or CT-A.  
 XX  
 SQ Sequence 103 AA;

Query Match 100.0%; Score 106; DB 17; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFQVEVPGSQHIDSQKAI 21  
 DB 45 gatfgvevpgsqhidsqkai 65

RESULT 4  
 AAW06606  
 ID AAW06606 standard; Protein; 103 AA.  
 AC AAW06606;

DT 06-AUG-1997 (first entry)

DE Cholera toxin B subunit/heat labile enterotoxin B subunit hybrid.

XX Heat labile; tolerant; stable; hybrid; cholera toxin B subunit; CTB;  
 KW heat labile enterotoxin B subunit; LTB; vaccine; immunisation;  
 KW immunogen; antigen; broad spectrum; cross reactive; diarrhoea;  
 KW vomiting; food poisoning.

XX Chimeric - Vibrio cholerae.  
 OS Chimeric - Enterotoxigenic Escherichia Coli.

XX Key Location/Qualifiers  
 FT Misc-difference 1  
 FT /label= substitution  
 FT /note= "wild-type Thr replaced by Ala"  
 FT Misc-difference 94  
 FT /label= substitution  
 FT /note= "wild-type His replaced by Asn"  
 FT Misc-difference 95  
 FT /label= substitution  
 FT /note= "wild-type Ala replaced by Ser"

XX WO9634893-Al.

XX 07-NOV-1996.

XX 02-MAY-1996; 96WO-SE00570.

XX 05-MAY-1995; 95SE-0001682.

XX (HOLM/) HOLMGREN J.  
 XX (LEBE/) LEBENS M R.

XX Holmgren J, Lebens MR;

XX WPI; 1996-506108/50.  
 XX N-PSDB; AAT43576.

XX Heat-labile enterotoxin B sub-unit and cholera toxin B sub-unit  
 PT hybrid protein - opt. fused to immunogenic sequence for use in  
 PT vaccines against enterotoxin-induced illness

XX Claim 3; Page -: 32pp; English.

XX AAW06606 is a mature cholera toxin B subunit (CTB)/heat labile  
 CC enterotoxin B subunit (LTB) hybrid molecule. In mature CTB protein,

CC certain amino acids (aa) were replaced with corresponding aa from  
 CC heat-labile enterotoxin B subunit (LTB). The specific amino acid  
 CC substitutions impart LTB-specific epitope characteristics to  
 CC immunogenic mature CTB. The hybrid molecules have increased  
 CC cross-reactivity and are suitable for a broad spectrum vaccine to  
 CC protect against enterotoxigenic illness. Immunogenic proteins  
 CC comprising the hybrid molecules can be used to treat, or in a  
 CC vaccine to prevent, enterotoxin-induced illness, e.g. diarrhoea  
 CC and vomiting, in humans and animals.  
 CC Note - This sequence does not appear in the specification, it is  
 CC a claimed mutant sequence of mature cholera toxin B subunit (see  
 CC AAW06605).

XX Sequence 103 AA;

Query Match 100.0%; Score 106; DB 17; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFQVEVPGSQHIDSQKAI 21  
 DB 45 gatfgvevpgsqhidsqkai 65

RESULT 5  
 AAW06607  
 ID AAW06607 standard; Protein; 103 AA.  
 AC AAW06607;

DT 06-AUG-1997 (first entry)

DE Cholera toxin B subunit/heat labile enterotoxin B subunit hybrid.

XX Heat labile; tolerant; stable; hybrid; cholera toxin B subunit; CTB;  
 KW heat labile enterotoxin B subunit; LTB; vaccine; immunisation;  
 KW immunogen; antigen; broad spectrum; cross reactive; diarrhoea;  
 KW vomiting; food poisoning.

XX Chimeric - Vibrio cholerae.  
 OS Chimeric - Enterotoxigenic Escherichia Coli.

XX Key Location/Qualifiers  
 FT Misc-difference 1.25  
 FT /label= substitution  
 FT /note= "the first 25 amino acids of mature  
 FT wild-type cholera toxin B subunit are  
 FT replaced with the first 25 amino acids  
 FT of mature enterotoxin B subunit"

XX WO9634893-Al.

XX 07-NOV-1996.

XX 02-MAY-1996; 96WO-SE00570.

XX 05-MAY-1995; 95SE-0001682.

XX (HOLM/) HOLMGREN J.  
 XX (LEBE/) LEBENS M R.

XX Holmgren J, Lebens MR;

XX WPI; 1996-506108/50.  
 XX N-PSDB; AAT43577.

XX Heat-labile enterotoxin B sub-unit and cholera toxin B sub-unit  
 PT hybrid protein - opt. fused to immunogenic sequence for use in  
 PT vaccines against enterotoxin-induced illness

XX Claim 4; Page -: 32pp; English.

AAW06607 is a mature cholera toxin B subunit (CTB)/heat labile enterotoxin B subunit (LTB) hybrid molecule. In mature CTB protein, certain amino acids (aa) were replaced with corresponding aa from heat-labile enterotoxin B subunit (LTB). The specific amino acid substitutions impart LTB-specific epitope characteristics to immunogenic mature CTB. The hybrid molecules have increased cross-reactivity and are suitable for a broad spectrum vaccine to protect against enterotoxigenic illness. Immunogenic proteins comprising the hybrid molecules can be used to treat, or in a vaccine to prevent, enterotoxin-induced illness, e.g. diarrhoea and vomiting, in humans and animals.

Note - This sequence does not appear in the specification, it is a claimed mutant sequence of mature cholera toxin B subunit (see AAW06605).

XX Sequence 103 AA;

Query Match 100.0%; Score 106; DB 17; Length 103;  
Best Local Similarity 100.0%; Pred. No. 1.7e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHDSQKKAI 21  
|||||  
DB 45 gatfqvevpgsqhidsqkkai 65

RESULT 6

AAW0808

ID AAW0808 standard; protein; 103 AA.

AC AAW0808;

XX

DT 29-JAN-1999 (first entry)

XX

DE Amino acid sequence of the wild type cholera toxin B subunit.

XX

KW Cholera toxin B subunit; nontoxic subunit; adjuvant; coadministration; antigen; bird; animal; mucosal; vaccine.

XX

OS Vibrio cholerae.

XX

PN W09845324-A1.

XX

PD 15-OCT-1998.

XX

PF 03-APR-1998; 98WO-US06725.

XX

PR 04-APR-1997; 97US-0043410.

XX

PA (KIYO/) KIYONO H.

PA (MCGH/) MCGHEE J. R.

PA (TAKE/) TAKEDA Y.

PA (JABR-) JAB RES FOUND.

PA (YAMA/) YAMAMOTO S.

XX

PI Kiyono H, Mcghee JR, Takeda Y, Yamamoto S;

XX

DR WPI; 1998-594478/50.

XX

PT New mutant cholera toxin selected from a group comprising nontoxic subunits/derivatives - effective as an adjuvant when coadministered with an antigen to birds and mammals

PT

XX

PS Disclosure; Fig 1B; 43pp; English.

XX

CC This is the amino acid sequence of the cholera toxin B subunit used in the method of the invention involving the use of nontoxic subunits as an effective adjuvant in coadministration of an antigen to birds and animals. In addition to the use of the toxin as an mucosal adjuvant, it also provides a vaccine comprising the toxin, an immunogenic amount of an antigen, and a pharmaceutically acceptable carrier. The toxin can be used with single/multiple vaccines, and it

CC enables the possibility for commercial mucosal adjuvants for use in humans, since these are more effective and safer than vaccines administered subcutaneously.

XX

SQ Sequence 103 AA;

Query Match 100.0%; Score 106; DB 19; Length 103;  
Best Local Similarity 100.0%; Pred. No. 1.7e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHDSQKKAI 21  
|||||  
DB 45 gatfqvevpgsqhidsqkkai 65

RESULT 7

AAW04163

ID AAW04163 standard; protein; 118 AA.

XX

AC AAW04163;

XX

DT 10-SEP-1990 (first entry)

XX

DE Cholera Toxin B-subunit.

XX

KW cholera toxin B-subunit; hybrid protein; heterologous IgA active antigen.

XX

OS synthetic.

XX

FH Key Location/Qualifiers

FT misc\_difference 18..18 /\*label= His or Tyr

FT 1..11

FT /\*label= signal peptide

FT /\*note= absent from mature protein

XX

PN W09003437-A.

XX

PD 05-APR-1990.

XX

PF 27-SEP-1989; 89WO-0000495.

XX

PR 27-SEP-1988; 88FR-0012627.

XX

PA (UYLI-) L'UNIVERSITE DE L'ETAT A LIEGE.

XX

PI L'Hoir C, Renard A, Martial J;

XX

DR WPI; 1990-132273/17.

XX

DR N-NSDB; Q04046.

XX

PT New hybrid protein, useful in vaccines - contains cholera toxin b subunit and heterologous IgA active antigenic sequence.

PT

XX

PS Disclosure; ; pp; French.

XX

CC Mature cholera toxin B-subunit is obtained when the signal peptide is cleaved off. There is an Ochre codon at position 343-5; the sequence downstream from it is part of a plasmid.

CC

XX

SQ Sequence 118 AA;

Query Match 100.0%; Score 106; DB 11; Length 118;  
Best Local Similarity 100.0%; Pred. No. 2e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHDSQKKAI 21  
|||||  
DB 56 gatfqvevpgsqhidsqkkai 76



## RESULT 8

AA093561  
 ID AAP93561 standard; protein; 124 AA.  
 XX AC  
 XX AAP93561;  
 DT 06-JUN-1990 (first entry)  
 XX  
 DE B subunit of the heat-labile enterotoxin (LT-B) derived from E. coli.  
 XX  
 KW B subunit; heat-labile enterotoxin; LT-B; Escherichia coli; malaria;  
 KW circumsporozoite protein; fusion protein; live recombinant vaccine;  
 KW Salmonella; epitope.  
 XX  
 OS Escherichia coli.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..22  
 FT /note="Signal peptide"  
 FT Protein 23..124  
 FT /note="Mature LT-B"  
 XX  
 XX W08902924-A.  
 XX  
 XX 06-APR-1989.  
 XX  
 XX 30-SEP-1988; 88WO-US03376.  
 XX  
 XX 02-OCT-1987; 87US-0104735.  
 XX  
 XX (PRAX-) PRAXIS BIOLOGICS IN.  
 XX  
 XX Brey RN, Majarian WR, Pillai S, Hockmeyer WT;  
 XX WPI; 1989-114399/15.  
 XX N-PSDB; AAN90747.  
 XX  
 XX Live recombinant vaccine for malaria.  
 XX comprising attenuated entero-invasive bacterium contg. DNA  
 XX encoding epitope of malaria parasite  
 XX  
 XX Fig 3; p. 3/17; 105pp; English.  
 XX  
 CC In the patent, the DNA encoding LR-B is expressed as part of a fusion  
 CC protein with an epitope of a malaria parasite, eg Region I or Region II  
 CC or a repeat region of circumsporozoite protein antigen (CS) (AAP93560)  
 CC from Plasmodium berghei. Pref. the fusion gene is inserted into  
 CC attenuated Salmonella enteritidis under the left promoter control of  
 CC lambda. Such bacteria can multiply in the host without causing disease or  
 CC disorder and express CS that will induce a protective immune response  
 CC against malaria and can be used in vaccines. Such vaccines can be  
 CC multivalent.  
 XX  
 SQ Sequence 124 AA;

Query Match 100.0%; Score 106; DB 10; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPCSHIDSQKKAI 21  
 |||||  
 Db 66 gatfvevpqshidsqkkai 86

## RESULT 9

AAW06605  
 ID AAW06605 standard; Protein; 124 AA.  
 XX AC  
 XX AAW06605;  
 DT 06-AUG-1997 (first entry)

XX Cholera toxin B subunit, used for hybrid immunogenic toxin production.  
 DE  
 XX  
 KW Heat labile; tolerant; stable; hybrid; cholera toxin B subunit; CTB;  
 KW heat labile enterotoxin B subunit; LT-B; vaccine; immunisation;  
 KW immunogen; antigen; broad spectrum; cross reactive; diarrhoea;  
 KW vomiting; food poisoning.  
 XX  
 OS Vibrio cholerae.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /label= sig\_peptide  
 FT Protein 22..124  
 FT /label= mat\_protein  
 XX  
 XX W09634893-A1.  
 XX  
 XX 07-NOV-1996.  
 XX  
 XX 02-MAY-1996; 96WO-SE00570.  
 XX  
 XX 05-MAY-1995; 95SE-0001682.  
 XX  
 XX (HOLM/) HOLMGREN J.  
 XX (LEBE/) LEBENS M R.  
 XX  
 XX Holmgren J, Lebens MR;  
 XX WPI; 1996-506108/50.  
 XX N-PSDB; AAT43575.  
 XX  
 XX Heat-labile enterotoxin B sub-unit and cholera toxin B sub-unit  
 XX hybrid protein - opt. fused to immunogenic sequence for use in  
 XX vaccines against enterotoxin-induced illness  
 XX  
 XX Disclosure; Fig 1; 32pp; English.  
 XX  
 CC AAW06605 is the full (including the signal peptide) length cholera  
 CC toxin B subunit (CTB), this sequence is described as unpublished in  
 CC the specification. The mature CTB protein was used to create hybrid  
 CC mutants, in which certain amino acids (aa) of CTB were replaced with  
 CC corresponding aa from heat-labile enterotoxin B subunit (LTB), see  
 CC AAW06606 and AAW06607. The specific amino acid substitutions impart  
 CC LTB-specific epitope characteristics to immunogenic mature CTB. The  
 CC hybrid molecules have increased cross-reactivity and are suitable  
 CC for a broad spectrum vaccine to protect against enterotoxigenic  
 CC illness. Immunogenic proteins comprising the hybrid molecules can be  
 CC used to treat, or in a vaccine to prevent, enterotoxin-induced illness,  
 CC e.g. diarrhoea and vomiting, in humans and animals.  
 XX  
 SQ Sequence 124 AA;

Query Match 100.0%; Score 106; DB 17; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPCSHIDSQKKAI 21  
 |||||  
 Db 66 gatfvevpqshidsqkkai 86

RESULT 10  
 AAW59770  
 ID AAW59770 standard; Protein; 124 AA.  
 XX AC  
 XX AAW59770;  
 XX  
 DT 12-OCT-1998 (first entry)  
 XX  
 DE Amino acid sequence of E. coli LTB.



```

FT Peptide 25..28
FT /label=part of CTB leader sequence
FT Protein 29..131
FT /label=mature CTB
FT Cleavage-site 24..25
FT /label=cleavage to release mature CTB
FT Cleavage-site 25..26
FT /label=cleavage to release mature CTB
FT Misc-difference 46
FT /label=H = Y in E1 Tor CTB
FT Misc-difference 75
FT /label=T = I in E1 Tor CTB
FT Misc-difference 82
FT /label=G = S in E1 Tor CTB
FT Misc-difference 50
FT /label=D = N in classical 569B CTB
FT Misc-difference 98
FT /label=D = N in classical 569 CTB
XX
XX EP368819-A.
XX
XX 16-MAY-1990.
XX
XX 11-SEP-1989; 89EP-0850295.
XX
XX 16-SEP-1988; 88SE-0003291.
XX 15-SEP-1989; 89NO-0003702.
XX
XX (HOLM/) HOLMGREN J.
XX
XX Holmgren J, Sanches C;
XX
XX WPI; 1990-149724/20.
XX P-PSDB; AAR04825.
XX
XX Expression of binding sub-unit protein of cholera toxin - using foreign
XX promoter with no V cholera DNA between promoter and ribosome binding site
XX
XX Disclosure; ; p; English.
XX
XX The sequence is produced by genetically fusing the leader sequence for
XX E.coli heat-labile enterotoxin subunit (LTB) by its 3' SacI end to the
XX 5' NdeI end of the cholera toxin subunit (CTB) via a synthetic linker.
XX This allows the use of a strategically placed EcoRI site just upstream
XX of the RBS on the LTB gene for the insertion of a strong tac promoter
XX for the expression of CTB. The protein can be used as vaccines,
XX diagnostic reagents and receptor-blocking agents for prophylaxis of
XX cholera and E.coli diarrhoea.
XX
XX Sequence 131 AA;

Query Match 100.0%; Score 106; DB 11; Length 131;
Best Local Similarity 100.0%; Pred. No. 2.3e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFQVEVPGSQHDSQKAI 21
DB 73 gatfqvevpgsqhdsqkai 93
|||||

RESULT 13
AAR50227
ID AAR50227 standard; Protein; 138 AA.
XX
XX AAR50227;
XX
XX 09-OCT-1994 (first entry)
XX
XX Sequence of LT-B-M24 hybrid molecule.
XX
XX B subunit; labile toxin; M protein; fusion protein; antigen;
XX Group A streptococci; rheumatic fever; pharyngitis.
KW

```

```

XX Synthetic.
XX WO9406465-A.
XX 31-MAR-1994.
XX
XX 15-SEP-1993; 93WO-US08704.
XX
XX 16-SEP-1992; 92US-0945860.
XX
XX (UYTE-) UNIV TENNESSEE RES CORP.
XX
XX Dale JB;
XX
XX WPI; 1994-118162/14.
XX N-PSDB; AAQ45159.
XX
XX New recombinant hybrid streptococcal M protein antigen(s) - which
XX elicit opsonic antibodies without eliciting cross-reactive
XX antibodies to mammalian heart tissue
XX
XX Example; Fig 1; 45pp; English.
XX
XX The surface M protein of Group A streptococci is the major virulence
XX factor and protective antigen of these organisms. However, there are
XX a tremendous number of M protein serotypes. The invention provides
XX recombinant M protein antigens comprising a gene encoding a carrier
XX protein and an NH2 or COOH terminal M protein fragment carrying one
XX or more epitopes. The carrier may be the B subunit of E.coli labile
XX toxin (LT-B). The carrier and the antigen may be linked by a linker,
XX eg AARS0226. The LT-B-M24 fusion gene of the example was expressed
XX using E.coli. The M24 component consists of a pair of synthetic
XX oligos which copied the first 36 bp of the emm 24 gene. Rabbits
XX immunised with the LT-B-M24 developed type-specific bacteriocidal
XX antibodies against type 24 streptococci.
XX
XX Sequence 138 AA;

Query Match 100.0%; Score 106; DB 15; Length 138;
Best Local Similarity 100.0%; Pred. No. 2.4e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFQVEVPGSQHDSQKAI 21
DB 66 gatfqvevpgsqhdsqkai 86
|||||

RESULT 14
AAW94082
ID AAW94082 standard; Protein; 170 AA.
XX
XX AAW94082;
XX
XX 09-APR-1999 (first entry)
XX
XX LT-B-CTP fusion protein.
XX
XX Self-tolerance; fusion protein; immune response; LTB; CTP; hCG; cancer;
XX labile toxin subunit B; carboxy terminal peptide; prophylaxis; fertility;
XX human chorionic gonadotrophin; LTB-CTP.
XX
XX Escherichia coli.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..21
XX /note= "signal peptide"
XX Protein 22..124
XX /note= "LTB protein"
XX Peptide 125..133
XX /note= "linker peptide"
XX

```



PI Sela M, Arnon R, Jacob CO;  
 DR WPI; 1985-069683/12.  
 XX Vaccines against cholera and heat-labile E. coli toxin - contg.  
 PT cholera toxin fragment coupled to carrier  
 XX Example; Fig 1; 24pp; German.  
 PS  
 XX The inventors claim vaccines against cholera and heat-labile E. coli  
 CC toxin contg. cholera toxin fragment coupled to carrier. The toxin is  
 CC esp. the fragments defined in FT, above.  
 XX  
 SQ Sequence 103 AA;

Query Match 97.2%; Score 103; DB 6; Length 103;  
 Best Local Similarity 95.2%; Pred. No. 5.6e-10;  
 Matches 20; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEPGSQHDSQKKAI 21  
 |||||:|||||:|||||:|||||  
 Db 45 gatfevpgsqhdsqkkai 65

RESULT 17  
 AAY87463  
 ID AAY87463 standard; peptide; 21 AA.  
 XX  
 AC AAY87463;  
 XX  
 DT 03-JUL-2000 (first entry)  
 XX  
 DE E. coli heat labile enterotoxin B-derived peptide, SEQ ID NO:5.  
 XX  
 KW Heat labile enterotoxin subunit B; EtxB;  
 KW beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;  
 KW adjuvant; immune disorder; diarrhoea.  
 XX  
 OS Escherichia coli.  
 XX  
 PN WO200014114-A1.  
 XX  
 PD 16-MAR-2000.  
 XX  
 PF 07-SEP-1999; 99WO-GB02970.  
 XX  
 PR 07-SEP-1998; 98GB-0019484.  
 XX  
 PA (UYBR-) UNIV BRISTOL.  
 XX  
 PI Williams NA, Hirst TR;  
 XX  
 DR WPI; 2000-256943/22.  
 XX  
 PT Derivatives of Escherichia coli heat labile enterotoxins useful as  
 PT immunomodulators and for treating diarrhoea and which do not bind the  
 PT glycolipid receptor GM-1 -  
 XX  
 PS Disclosure; Page 15; 62pp; English.  
 XX  
 CC The invention relates to peptide fragments of the Escherichia coli heat  
 CC labile enterotoxin (Etx) and its closely related homologue, cholera  
 CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous  
 CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are  
 CC composed of one A subunit and five identical B subunits. The A subunit  
 CC is responsible for toxicity, possessing adenosine diphosphate (ADP)-  
 CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)  
 CC facilitate the entry of subunit A into the host cell via the binding and  
 CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible  
 CC for some of the effects of Etx and Ctx, it has been found that certain  
 CC effects of the toxins, such as immunomodulation, are not mediated  
 CC through GM-1 binding. The peptides of the invention are fragments of the

CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as  
 CC normal EtxB and CtxB subunits, except that they do not bind or cross link  
 CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.  
 CC Therefore, the peptides may be used in the production of a composition  
 CC for treating, preventing and/or modulating a disease associated with an  
 CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463  
 CC represent preferred peptides of the invention, AAY87460 being  
 CC particularly preferred.  
 XX  
 SQ Sequence 21 AA;

Query Match 95.3%; Score 101; DB 21; Length 21;  
 Best Local Similarity 95.2%; Pred. No. 1.9e-10;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATFOVEPGSQHDSQKKAI 21  
 |||||:|||||:|||||:|||||  
 Db 1 getfqvepgsqhdsqkkai 21

RESULT 18  
 AAR72545  
 ID AAR72545 standard; peptide; 93 AA.  
 XX  
 AC AAR72545;  
 XX  
 DT 28-NOV-1995 (first entry)  
 XX  
 DE ADP-ribosylating toxin (verotoxin-1 B-subunit).  
 XX  
 KW ADP-ribosylating toxin; pertussis holotoxin; B-subunit;  
 KW active site; E. coli heat labile toxin; verotoxin-1;  
 KW Bordetella pertussis vaccines.  
 XX  
 OS Bacteria sp.  
 XX  
 PN EP646599-A.  
 XX  
 PD 05-APR-1995.  
 XX  
 PF 23-AUG-1994; 94EP-0306219.  
 XX  
 PR 24-AUG-1993; 93US-0110947.  
 XX  
 PR 31-MAY-1994; 94US-0251121.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 PA (UYAL-) UNIV ALBERTA.  
 XX  
 PI Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;  
 PI Oomen R, Read RJ, Stein PE;  
 XX  
 DR WPI; 1995-132623/18.  
 XX  
 PT New modified forms of pertussis holotoxin - developed using  
 PT crystalline forms of pertussis holotoxin and its complexes with  
 PT other molecules  
 XX  
 PS Disclosure; Fig 5; 54pp; English.  
 XX  
 CC AAR72540-R72545 are structurally equivalent B-subunits from three  
 CC ADP-ribosylating toxins, pertussis holotoxin (PT), E. coli heat  
 CC labile toxin (LT), and verotoxin-1 (VT). The structural  
 CC information obtd. from these comparisons was used to identify  
 CC sites which contribute to PT's biological activity. By modifying  
 CC these sites the claimed PT mutants of the invention were produced,  
 CC they can be used in the development of vaccines against Bordetella  
 CC pertussis infection.  
 XX  
 SQ Sequence 93 AA;



RESULT 21  
AAY68365  
ID AAY68365 standard; Peptide; 93 AA.  
XX  
AC AAY68365;  
XX  
DT 17-APR-2000 (first entry)  
XX  
DE Heat labile toxin B subunit SEQ ID NO:26.  
XX  
KW Pertussis toxin; heat labile toxin; verotoxin; exotoxin A; holotoxin;  
KW diphtheria toxin; ADP-ribosylating toxin; mannose binding protein;  
KW infection; crystal structure; X-ray crystallography; detoxification;  
KW immunogenic.  
XX  
OS Escherichia coli.  
XX  
PN US6018022-A.  
XX  
XX 25-JAN-2000.  
XX  
XX 06-JUN-1995; 95US-0467976.  
XX  
XX 22-AUG-1994; 94US-0292968.  
XX  
XX 24-AUG-1993; 93US-0110947.  
XX  
XX 31-MAY-1994; 94US-0251121.  
XX  
XX (CONN-) CONNAUGHT LAB LTD.  
XX  
XX (UNAL-) UNIV ALBERTA.  
XX  
XX Read RJ, Stein PE, Cockle SA, Loosmore S, Klein MH, Armstrong GD;  
XX Haze B, Oomen RP;  
XX  
XX WPI; 2000-136703/12.  
XX  
XX Crystalline form of isolated pertussis holotoxin useful in studying  
XX proteins which have functional resemblance -  
XX  
XX Example 3; Fig 5; 42pp; English.  
XX  
XX The present invention describes a crystalline form of isolated  
XX pertussis holotoxin, in which the molecules of pertussis toxin have  
XX a three dimensional structure represented in the specification,  
XX complexed with a polysaccharide molecule capable of forming a complex  
XX with the holotoxin. The crystalline form of the pertussis holotoxin  
XX can be used in a comparison with other proteins which have functional  
XX resemblance to pertussis holotoxin with the aim of modifying other  
XX proteins. Identifying the unknown sites of toxicity by comparison  
XX with the three dimensional structure of pertussis holotoxin provides a  
XX technique for detoxification of toxins to produce useful immunogenic  
XX but non-toxic analogues. It can also be used as a primary standard to  
XX measure the quantity, purity or efficacy of less pure compositions  
XX containing pertussis toxin. AAY68340 to AAY68385 represent peptides  
XX used in the exemplification of the present invention.  
XX  
XX Sequence 93 AA;  
XX  
Query Match 95.3%; Score 101; DB 21; Length 93;  
Best Local Similarity 95.2%; Pred. No. 1.1e-09;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GATFQVEVPGSQHDSQKKAI 21  
Db 35 getfqvevpgsqhdsqkkai 55  
XX  
RESULT 22  
AAB66239  
ID AAB66239 standard; Protein; 93 AA.  
XX

AAB66239;  
XX  
XX 03-APR-2001 (first entry)  
XX  
DE E coli verotoxin-1 B subunit SEQ ID NO: 26.  
XX  
XX Pertussis toxin; crystal structure; whooping cough; biological activity;  
KW lymphocytosis-promoting factor; histamine-sensitising factor;  
KW islet-activating protein.  
XX  
XX Escherichia coli.  
XX  
XX US6168928-B1.  
XX  
XX 02-JAN-2001.  
XX  
XX 21-MAY-1998; 98US-0082514.  
XX  
XX 22-AUG-1994; 94US-0292968.  
XX  
XX 24-AUG-1993; 93US-0110947.  
XX  
XX 31-MAY-1994; 94US-0251121.  
XX  
XX (CONN-) CONNAUGHT LAB LTD.  
XX  
XX Read RJ, Cockle SA, Oomen RP, Loosmore S, Klein MH, Armstrong GD;  
XX Haze B, Stein PE;  
XX  
XX WPI; 2001-122260/13.  
XX  
XX Modifying pertussis holotoxin to produce detoxified PT analogs,  
XX comprising analyzing crystalline structure of toxin, to identify sites  
XX of toxicity, cell binding or enzyme activity of PT and modifying  
XX identified site -  
XX  
XX Example 3; Fig 5; 4lpp; English.  
XX  
XX The present invention provides a method for producing a pertussis toxin  
XX (also designated lymphocytosis-promoting factor, histamine-sensitising  
XX factor and islet activating protein) with a modified biological activity,  
XX involving analysing the crystal structure of the protein to identify  
XX active sites which can then be modified. This may lead to an alteration  
XX in the toxicity, cell binding or enzyme activity of the toxin. This can  
XX be used in the production of immunoprotective analogues of pertussis  
XX toxin. Pertussis toxin is the cause of whooping cough following infection  
XX by Bordetella pertussis.  
XX  
XX Sequence 93 AA;  
XX  
Query Match 95.3%; Score 101; DB 22; Length 93;  
Best Local Similarity 95.2%; Pred. No. 1.1e-09;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GATFQVEVPGSQHDSQKKAI 21  
Db 35 getfqvevpgsqhdsqkkai 55  
XX  
RESULT 23  
AAB73241  
ID AAB73241 standard; Protein; 134 AA.  
XX  
XX AAB73241;  
XX  
XX 14-MAY-2001 (first entry)  
XX  
XX Recombinant exotoxin protein variant LTBpL.  
XX  
XX Exotoxin mucosal cell binding motif; nucleic acid delivery;  
KW nucleic acid affinity domain; heat-labile enterotoxin.  
XX  
XX Unidentified.

PN WO200111960-A1.  
XX  
PD 22-FEB-2001.  
XX  
PF 18-AUG-2000; 2000WO-US22715.  
XX  
PR 18-AUG-1999; 99US-0149294.  
XX  
PA (AGRI-) AGRIVAX INC.  
XX  
PI Welter LM;  
XX  
DR WPI; 2001-211103/21.  
XX  
DR N-PSDB; AAF75712.  
XX  
PT Novel exotoxin protein variant useful as protein carrier for  
PT facilitating gene delivery, comprises a mucosal cell binding motif of  
PT an exotoxin and a nucleic acid affinity domain  
XX  
PS Example 1; Fig 9; 57pp; English.  
XX  
CC The present invention relates to recombinant exotoxin protein variants,  
CC which comprise an exotoxin mucosal cell binding motif and a nucleic acid  
CC affinity domain. The present sequence is one such protein variant. In the  
CC present invention the heat-labile enterotoxin (LT) of Escherichia coli  
CC was used as an exotoxin. LT is a member of the ADP-ribosylating exotoxin  
CC (bare) family. The protein variants are useful for selectively delivering  
CC nucleic acid to mucosal cells, for inducing an immune response when the  
CC nucleic acid encodes an antigen to which the immune response is desired,  
CC for selectively delivering a gene to a mucosal cell, and for achieving  
CC expression of a protein in a subject, by administering a composition  
CC comprising the protein variant.  
XX  
SQ Sequence 134 AA;  
  
Query Match 95.3%; Score 101; DB 22; Length 134;  
Best Local Similarity 95.2%; Pred. No. 1.7e-09;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 GATFOVEVPGSQHIDSOKKAI 21  
| | | | | | | | | | | | | | | | | | | | | |  
Db 66 getfqvevpqsgnidsqkka1 86  
  
RESULT 24  
AAB73242  
ID AAB73242 standard; Protein; 142 AA.  
XX  
AC AAB73242;  
XX  
DT 14-MAY-2001 (first entry)  
XX  
DE Recombinant exotoxin protein variant LTBpLh.  
XX  
KW Exotoxin mucosal cell binding motif; nucleic acid delivery;  
KW nucleic acid affinity domain; heat-labile enterotoxin.  
XX  
OS Unidentified.  
XX  
PN WO200111960-A1.  
XX  
PD 22-FEB-2001.  
XX  
PF 18-AUG-2000; 2000WO-US22715.  
XX  
PR 18-AUG-1999; 99US-0149294.  
XX  
PA (AGRI-) AGRIVAX INC.  
XX  
PI Welter LM;  
XX  
DR WPI; 2001-211103/21.  
XX  
PT Novel exotoxin protein variant useful as protein carrier for  
PT facilitating gene delivery, comprises a mucosal cell binding motif of  
PT an exotoxin and a nucleic acid affinity domain  
XX  
PS Example 1; Fig 11; 57pp; English.  
XX  
CC The present invention relates to recombinant exotoxin protein variants,  
CC which comprise an exotoxin mucosal cell binding motif and a nucleic acid  
CC affinity domain. The present sequence is one such protein variant. In the  
CC present invention the heat-labile enterotoxin (LT) of Escherichia coli  
CC was used as an exotoxin. LT is a member of the ADP-ribosylating exotoxin  
CC (bare) family. The protein variants are useful for selectively delivering  
CC nucleic acid to mucosal cells, for inducing an immune response when the  
CC nucleic acid encodes an antigen to which the immune response is desired,  
CC for selectively delivering a gene to a mucosal cell, and for achieving  
CC expression of a protein in a subject, by administering a composition  
CC comprising the protein variant.  
XX  
SQ Sequence 134 AA;

DR N-PSDB; AAF75713.  
XX  
PT Novel exotoxin protein variant useful as protein carrier for  
PT facilitating gene delivery, comprises a mucosal cell binding motif of  
PT an exotoxin and a nucleic acid affinity domain  
XX  
PS Example 1; Fig 10; 57pp; English.  
XX  
CC The present invention relates to recombinant exotoxin protein variants,  
CC which comprise an exotoxin mucosal cell binding motif and a nucleic acid  
CC affinity domain. The present sequence is one such protein variant. In the  
CC present invention the heat-labile enterotoxin (LT) of Escherichia coli  
CC was used as an exotoxin. LT is a member of the ADP-ribosylating exotoxin  
CC (bare) family. The protein variants are useful for selectively delivering  
CC nucleic acid to mucosal cells, for inducing an immune response when the  
CC nucleic acid encodes an antigen to which the immune response is desired,  
CC for selectively delivering a gene to a mucosal cell, and for achieving  
CC expression of a protein in a subject, by administering a composition  
CC comprising the protein variant.  
XX  
SQ Sequence 142 AA;  
  
Query Match 95.3%; Score 101; DB 22; Length 142;  
Best Local Similarity 95.2%; Pred. No. 1.8e-09;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 GATFOVEVPGSQHIDSOKKAI 21  
| | | | | | | | | | | | | | | | | | | | | |  
Db 66 getfqvevpqsgnidsqkka1 86  
  
RESULT 25  
AAB73243  
ID AAB73243 standard; Protein; 155 AA.  
XX  
AC AAB73243;  
XX  
DT 14-MAY-2001 (first entry)  
XX  
DE Recombinant exotoxin protein variant LTB-P.  
XX  
KW Exotoxin mucosal cell binding motif; nucleic acid delivery;  
KW nucleic acid affinity domain; heat-labile enterotoxin.  
XX  
OS Unidentified.  
XX  
PN WO200111960-A1.  
XX  
PD 22-FEB-2001.  
XX  
PF 18-AUG-2000; 2000WO-US22715.  
XX  
PR 18-AUG-1999; 99US-0149294.  
XX  
PA (AGRI-) AGRIVAX INC.  
XX  
PI Welter LM;  
XX  
DR WPI; 2001-211103/21.  
XX  
DR N-PSDB; AAF75714.  
XX  
PT Novel exotoxin protein variant useful as protein carrier for  
PT facilitating gene delivery, comprises a mucosal cell binding motif of  
PT an exotoxin and a nucleic acid affinity domain  
XX  
PS Example 1; Fig 11; 57pp; English.  
XX  
CC The present invention relates to recombinant exotoxin protein variants,  
CC which comprise an exotoxin mucosal cell binding motif and a nucleic acid  
CC affinity domain. The present sequence is one such protein variant. In the  
CC present invention the heat-labile enterotoxin (LT) of Escherichia coli  
CC was used as an exotoxin. LT is a member of the ADP-ribosylating exotoxin



AA  
SQ  
Sequence  
155 AA;

**Qy**    1 GATFQVEVPGSQHIDSQKKAI 21  
         | |||||  
**Db**    66 getfqvevpgsqhidsqkkai 86

Search completed: July 16, 2001, 16:35:41  
Job time: 205 sec

***This Page Blank (uspto)***

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: July 16, 2001, 16:36:18 ; Search time 30.3 seconds  
(without alignments)  
13.962 Million cell updates/sec

Title: US-09-786-648-4  
Perfect score: 106  
Sequence: 1 GATQVEVPGSQHDSQKAI 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	100.0	102	3	US-08-952-337-5
2	106	100.0	102	3	US-08-952-337-6
3	106	100.0	103	2	US-08-472-171-2
4	106	100.0	103	2	US-08-894-526-2
5	106	100.0	103	2	US-09-013-047-2
6	106	100.0	103	4	US-09-374-597-2
7	106	100.0	103	4	US-09-191-852-21
8	106	100.0	103	5	PCT-US95-13376-21
9	106	100.0	123	3	US-08-952-337-1
10	106	100.0	123	3	US-08-952-337-2
11	106	100.0	124	2	US-08-747-410-2
12	101	95.3	93	2	US-08-292-968-26
13	101	95.3	93	2	US-08-467-974-26
14	101	95.3	93	2	US-08-467-536-26
15	101	95.3	93	3	US-08-467-976-26
16	101	95.3	93	3	US-09-082-514-26
17	101	95.3	371	2	US-08-829-026A-6
18	94	88.7	124	1	US-08-449-045C-4
19	94	88.7	124	1	US-08-435-605A-12
20	94	88.7	124	6	5223610-3
21	51	48.1	448	2	US-08-878-989-2
22	51	48.1	448	4	US-09-272-796-2
23	42	39.6	855	2	US-09-027-337-2
24	40.5	38.2	862	3	US-08-709-784-2
25	40.5	38.2	862	2	US-08-209-521-23
26	40.5	38.2	862	2	US-08-209-521-30
27	40.5	38.2	862	4	US-09-059-461-2

28	40.5	38.2	862	4	US-08-961-810-133	Sequence 133, App
29	40.5	38.2	862	4	US-08-352-902B-133	Sequence 133, App
30	39	36.8	251	1	US-07-956-700B-94	Sequence 94, Appl
31	39	36.8	251	1	US-08-476-537-94	Sequence 94, Appl
32	39	36.8	251	1	US-08-485-607-94	Sequence 94, Appl
33	39	36.8	251	2	US-08-475-879-94	Sequence 94, Appl
34	39	36.8	346	2	US-08-602-359A-34	Sequence 34, Appl
35	39	36.8	384	4	US-09-025-580-35	Sequence 35, Appl
36	39	36.8	427	4	US-09-025-580-36	Sequence 36, Appl
37	39	36.8	459	6	5194375-6	Patent No. 5194375
38	39	36.8	775	2	US-08-714-070A-1	Sequence 1, Appl
39	38	35.8	845	1	US-08-416-950-11	Sequence 11, Appl
40	38	35.8	845	2	US-08-469-830-11	Sequence 11, Appl
41	38	35.8	1810	5	PCT-US95-11684-4	Sequence 4, Appl
42	37	34.9	75	1	US-08-350-884-35	Sequence 35, Appl
43	37	34.9	75	1	US-08-709-173-35	Sequence 35, Appl
44	37	34.9	75	2	US-08-709-177-35	Sequence 35, Appl
45	37	34.9	75	2	US-08-833-678A-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1  
US-08-952-337-5  
; Sequence 5, Application US/08952337  
; Patent No. 6019973  
; GENERAL INFORMATION:  
; APPLICANT: Holmgren, Jan  
; APPLICANT: Lebens, Michael R.  
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE  
; TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS  
; FILE REFERENCE: 3846/OD758  
; CURRENT APPLICATION NUMBER: US/08/952,337  
; CURRENT FILING DATE: 1998-01-05  
; EARLIER APPLICATION NUMBER: PCT/SE96/00570  
; EARLIER FILING DATE: 1996-05-02  
; EARLIER APPLICATION NUMBER: SE 9501682-0  
; EARLIER FILING DATE: 1995-05-05  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 102  
; TYPE: PRT  
; ORGANISM: Vibrio cholerae

US-08-952-337-5

Query Match 100.0%; Score 106; DB 3; Length 102;  
Best Local Similarity 100.0%; Pred. No. 2e-11; Indels 0; Gaps 0;  
Matches 21; Conservative 0; Mismatches 0

Qy 1 GATQVEVPGSQHDSQKAI 21  
|||||  
Db 44 GATQVEVPGSQHDSQKAI 64

RESULT 2  
US-08-952-337-6  
; Sequence 6, Application US/08952337  
; Patent No. 6019973  
; GENERAL INFORMATION:  
; APPLICANT: Holmgren, Jan  
; APPLICANT: Lebens, Michael R.  
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE  
; TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS  
; FILE REFERENCE: 3846/OD758  
; CURRENT APPLICATION NUMBER: US/08/952,337  
; CURRENT FILING DATE: 1998-01-05  
; EARLIER APPLICATION NUMBER: PCT/SE96/00570  
; EARLIER FILING DATE: 1996-05-02  
; EARLIER APPLICATION NUMBER: SE 9501682-0  
; EARLIER FILING DATE: 1995-05-05

; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 102  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-08-952-337-6

Query Match 100.0%; Score 106; DB 3; Length 102;  
Best Local Similarity 100.0%; Pred. No. 2e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHDSOKKAI 21  
Db 44 GATFOVEVPGSQHDSOKKAI 64

RESULT 3  
US-08-472-171-2  
; Sequence 2, Application US/08472171  
; Patent No. 5932714  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yacoub, Reza K.  
; APPLICANT: Zealey, Gavin R.  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Expression of Gene Products From  
; TITLE OF INVENTION: Genetically Manipulated Strains of Bordetella  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 330 University Avenue, Suite 701  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,171  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/393,334  
FILING DATE: 23-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-507 MIS:vg  
TELEPHONE: 416-595-1155  
TELEFAX: 416-595-1163  
TELEX: 065-24567 Simbas  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-472-171-2  
Query Match 100.0%; Score 106; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 2.1e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GATFOVEVPGSQHDSOKKAI 21  
Db 45 GATFOVEVPGSQHDSOKKAI 65

RESULT 3  
US-08-472-171-2  
; Sequence 2, Application US/08472171  
; Patent No. 5932714  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yacoub, Reza K.  
; APPLICANT: Zealey, Gavin R.  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Expression of Gene Products From  
; TITLE OF INVENTION: Genetically Manipulated Strains of Bordetella  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 330 University Avenue, Suite 701  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7

RESULT 4  
US-08-894-526-2  
; Sequence 2, Application US/08894526  
; Patent No. 5942418  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yacoub, Reza K.  
; APPLICANT: Zealey, Gavin R.  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: EXPRESSION OF GENE PRODUCTS FROM  
; TITLE OF INVENTION: GENETICALLY MANIPULATED STRAINS OF BORDETELLA  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/894,526  
FILING DATE: 01-DEC-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-724 MIS:jb  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-894-526-2

Query Match 100.0%; Score 106; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 2.1e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GATFOVEVPGSQHDSOKKAI 21  
Db 45 GATFOVEVPGSQHDSOKKAI 65

RESULT 5  
US-09-013-047-2  
; Sequence 2, Application US/09013047  
; Patent No. 5998168  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yacoub, Reza K.  
; APPLICANT: Zealey, Gavin R.  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Expression of Gene Products From  
; TITLE OF INVENTION: Genetically Manipulated Strains of Bordetella  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 330 University Avenue, 6th Floor  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada

```

; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/013,047
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/472,171
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/393,334
; FILING DATE: 23-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-507 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 416-595-1155
; TELEFAX: 416-595-1163
; TELEX: 065-24567 Simbas
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-013-047-2

Query Match 100.0%; Score 106; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.1e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHDSQKKAI 21
Db 45 GATFOVEVPGSQHDSQKKAI 65

RESULT 6
US-09-374-597-2
; Sequence 2, Application US/09374597
; Patent No. 6140082
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yacoub, Reza K.
; APPLICANT: Zealey, Gavin R.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Expression Of Gene Products From
; TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 330 University Avenue, 6th Floor
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/374,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/393,334
; FILING DATE: FEBRUARY 23, 1995

```

```

; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 416-595-1155
; TELEFAX: 416-595-1163
; TELEX: 065-24567 Simbas
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-374-597-2

Query Match 100.0%; Score 106; DB 4; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.1e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHDSQKKAI 21
Db 45 GATFOVEVPGSQHDSQKKAI 65

RESULT 7
US-09-191-852-21
; Sequence 21, Application US/09191852
; Patent No. 6194560
; GENERAL INFORMATION:
; APPLICANT: Charles J. Arntzen, Hugh S. Mason, and Tariq A. Haq
; TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski L.L.P.
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/191,852
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13376
; FILING DATE: 24-OCT-1995
; APPLICATION NUMBER: 08/817,906
; FILING DATE: 04-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Fox, David L.
; REGISTRATION NUMBER: 40,612
; REFERENCE/DOCKET NUMBER: P01590US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-651-5151
; TELEFAX: 713-651-5246
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-09-191-852-21

Query Match 100.0%; Score 106; DB 4; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.1e-11;

```

Matches	21;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	----	--------	----	------	----

Qy 1 GATFQVEVPGSQHIDSQKKAI 21  
          |||||  
Db 45 GATFQVEVPGSQHIDSQKKAI 65

Db 45 GATFOVEVPGSOHIDSOKKAI 65

## RESULT

```

PCT-US95-13376-21
; Sequence 21, Application PC/TUS9513376
;
; GENERAL INFORMATION:
;
; APPLICANT: The Texas A&M University System
;
; APPLICANT: 310 Wisenbaker
;
; APPLICANT: College Station, Texas 77843-3369
;
; TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS

```

```

Query Match      100.0%; Score 106; DB 5; Length 103;
Best Local Similarity 100.0%; Pred. NO. 2.1e-11;
Matches 21; Conservative 0; Mismatches 0; Indels

```

QY 1 GATFQVEVPGSQHIDSQKKAI 21  
|||  
db 45 GATFQVEVPGSQHIDSQKKAI 65

45 GATFOVEVPGSOHTDSOKKAT 65

## RESULTS

RESUB 9  
US-08-952-337-1  
; Sequence 1, Application US/08952337  
; Patent No. 6019973  
; GENERAL INFORMATION:  
; APPLICANT: Holmgren, Jan  
; APPLICANT: Lebens, Michael R.  
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE  
; TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS  
; FILE REFERENCE: 3846/0D758  
; CURRENT APPLICATION NUMBER: US/08/952,337  
; CURRENT FILING DATE: 1998-01-05  
; EARLIER APPLICATION NUMBER: PCT/SE96/00570  
; EARLIER FILING DATE: 1996-05-02

```

; EARLIER APPLICATION NUMBER: SE 9501682-0
; EARLIER FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 123
; TYPE: prt
; ORGANISM: Vibrio cholerae
US-08-952-337-1

```

```
Query Match      100.0%; Score 106; DB 3; Length 123;
Best Local Similarity 100.0%; Pred. No. 2.6e-11;
Matches 21; Conservative 0; Mismatches 0; Indels
```

QY 1 GATFQVEVPGSQHIDSQKKAI 21  
 |||||  
 Db 65 GATFOVEVPGSOHIDSOKKAI 85

db : 65 GATFOVEVPGSOHIDSOKKAI 85

RESULT 10

```

US-08-952-337-2
; Sequence 2, Application US/08952337
; Patent No. 6019973
; GENERAL INFORMATION:
; APPLICANT: Holmgren, Jan
; APPLICANT: Lebens, Michael R.
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
; TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS
; FILE REFERENCE: 3846/0D758
; CURRENT APPLICATION NUMBER: US/08/952,337
; CURRENT FILING DATE: 1998-01-05
; EARLIER APPLICATION NUMBER: PCT/SE96/00570
; EARLIER FILING DATE: 1996-05-02
; EARLIER APPLICATION NUMBER: SE 9501682-0
; EARLIER FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-952-337-2

```

```
Query Match      100.0%; Score 106; DB 3; Length 123;
Best Local Similarity 100.0%; Pred. No. 2.6e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0;
```

Qy 1 GATFQVEVPGSQHIDSQKKAI 21  
|||||  
Db 65 GATFOVEVPGSOHIDSQKKAI 85

65 GATFOVEVPGSSQHTDSQKKAT 85

11 JUL 1953

US-08-747-410-2  
 ; Sequence 2, Application US/08747410  
 ; Patent No. 5993820  
 ; GENERAL INFORMATION:  
 ; APPLICANT: BAGDASARIAN, Michael  
 ; APPLICANT: IRELAND, James  
 ; TITLE OF INVENTION: CHIMERIC LTB VACCINES  
 ; NUMBER OF SEQUENCES: 12  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt  
 ; STREET: 3100 N. 5993820west Center, 90 South Seventh St  
 ; CITY: Minneapolis  
 ; STATE: MN  
 ; COUNTRY: USA  
 ; ZIP: 55402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible

```
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/747,410
FILING DATE: 12-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Kettelberger, Denise M
REGISTRATION NUMBER: 33,924
REFERENCE/DOCKET NUMBER: 11526.1-US-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612/371-5268
TELEFAX: 612/332-9081
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
US-08-747-410-2
```

```
Query Match 100.0%; Score 106; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 2.6e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GATFOVEVPGSQHDSQKAI 21
|||||
Db 66 GATFOVEVPGSQHDSQKAI 86
```

```
RESULT 12
US-08-292-968-26
Sequence 26, Application US/08292968
Patent No. 5856122
GENERAL INFORMATION:
APPLICANT: READ, Randy J.
APPLICANT: STEIN, Penelope E.
APPLICANT: COCKLE, Stephen A.
APPLICANT: OOMEN, Raymond P.
APPLICANT: LOOSMORE, Sheena
APPLICANT: KLEIN, Michel H.
APPLICANT: ARMSTRONG, Glen D.
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/292,968
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/110,947
```

```
FILING DATE: 24-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/251,121
FILING DATE: 31-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-388
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-292-968-26
```

```
Query Match 95.3%; Score 101; DB 2; Length 93;
Best Local Similarity 95.2%; Pred. No. 1.4e-10;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GATFOVEVPGSQHDSQKAI 21
|||||
Db 35 GATFOVEVPGSQHDSQKAI 55
```

```
RESULT 13
US-08-467-974-26
Sequence 26, Application US/08467974
Patent No. 5965385
GENERAL INFORMATION:
APPLICANT: READ, Randy J.
APPLICANT: STEIN, Penelope E.
APPLICANT: COCKLE, Stephen A.
APPLICANT: OOMEN, Raymond P.
APPLICANT: LOOSMORE, Sheena
APPLICANT: KLEIN, Michel H.
APPLICANT: ARMSTRONG, Glen D.
APPLICANT: HAZES, Bart
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,974
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/467,536
FILING DATE: 06-JUN-1995
APPLICATION NUMBER: US 08/292,968
FILING DATE: 22-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/251,121
FILING DATE: 31-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/110,947
FILING DATE: 24-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
```

REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-454 MIS:vg  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-467-974-26

Query Match 95.3%; Score 101; DB 2; Length 93;  
Best Local Similarity 95.2%; Pred. No. 1.4e-10;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSOHIDSOKKAI 21  
Db 35 GETFOVEVPGSOHIDSOKKAI 55

RESULT 14  
US-08-467-536-26  
Sequence 26, Application US/08467536  
Patent No. 5977304  
GENERAL INFORMATION:  
APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,536  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/292,968  
FILING DATE: 22-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/251,121  
FILING DATE: 31-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-455 MIS:vg  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids

TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-467-536-26  
Query Match 95.3%; Score 101; DB 2; Length 93;  
Best Local Similarity 95.2%; Pred. No. 1.4e-10;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSOHIDSOKKAI 21  
Db 35 GETFOVEVPGSOHIDSOKKAI 55

RESULT 15  
US-08-467-976-26  
Sequence 26, Application US/08467976  
Patent No. 6018022  
GENERAL INFORMATION:  
APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,976  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/292,968  
FILING DATE: 22-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/251,121  
FILING DATE: 31-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-453 MIS:vg  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-467-976-26

Query Match 95.3%; Score 101; DB 3; Length 93;  
Best Local Similarity 95.2%; Pred. No. 1.4e-10;



Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATFQVEVPGSQHIDSOKKAI 21  
| | | | | | | | | | | | | | |  
Db 35 GETFQVEVPGSQHIDSOKKAI 55

RESULT 16  
US-09-082-514-26  
; Sequence 26, Application US/09082514  
; Patent No. 6168928  
; GENERAL INFORMATION:  
; APPLICANT: READ, Randy J.  
; APPLICANT: STEIN, Penelope E.  
; APPLICANT: COCKLE, Stephen A.  
; APPLICANT: COMEN, Raymond P.  
; APPLICANT: KLEIN, Michel H.  
; APPLICANT: ARMSTRONG, Glen D.  
; APPLICANT: HAZES, Bart  
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Slim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/082,514  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/292,968  
; FILING DATE: 24-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-810  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 93 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-082-514-26

Query Match 95.3%; Score 101; DB 4; Length 93;  
Best Local Similarity 95.2%; Pred. No. 1.4e-10;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATFQVEVPGSQHIDSOKKAI 21  
| | | | | | | | | | | | | | |  
Db 35 GETFQVEVPGSQHIDSOKKAI 55

RESULT 17  
US-08-829-026A-6  
; Sequence 6, Application US/08829026A  
; Patent No. 5837825  
; GENERAL INFORMATION:  
; APPLICANT: Meinersmann, Richard J.  
; APPLICANT: Khoury, Christian A.  
; TITLE OF INVENTION: Campylobacter Jejunii Flagellip-Escherichia Coli LT-B Fusion

```

; ; FILING DATE: 06-JUL-1994
; ; ATTORNEY/AGENCY INFORMATION:
; ; NAME: Mazza, Richard J.
; ; REGISTRATION NUMBER: 27,657
; ; REFERENCE/DOCKET NUMBER: A-196C
; ; INFORMATION FOR SEQ ID NO: 4:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 124 amino acids
; ; TYPE: amino acid
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: protein
; ; US-08-449-045C-4

```

Query Match 88.7%; Score 94; DB 1; Length 124;  
Best Local Similarity 90.5%; Pred. No. 3.3e-09;  
Matches 19; Conservative 0; Mismatches 2; Indels

QY      1 GATFQVEVPGSQHIDSQKAI 21  
         || ||||| |||||  
Db     66 GAIFQVEVPSSOHIDSQKAI 86

```

RESULT 19
US-08-435-605A-12
; Sequence 12, Application US/08435605A
; Patent No. 5874287
; GENERAL INFORMATION:
; APPLICANT: Burnette, W. Neal
; APPLICANT: Kaslow, Harvey R.
; TITLE OF INVENTION: RECOMBINANT DNA-DERIVED CHOLERA TOXIN
; TITLE OF INVENTION: SUBUNIT ANALOGS
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/435,605A
; FILING DATE: 05-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mazza, Richard J.
; REGISTRATION NUMBER: 27,657
; REFERENCE/POCKET NUMBER: A-196B
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 124 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-435-605A-12

```

Query Match 88.7%; Score 94; DB 2; Length 124;  
Best Local Similarity 90.5%; Pred. No. 3.3e-09;  
Matches 19; Conservative 0; Mismatches 2; Indels

QY	1	GATFQVEVP	GSQHIDSQK	KAI	21
db	66	GAIFQVEVP	SSOHIDSQK	KAI	86

RESULT 20

5223610-3  
 ; Patent No. 5223610  
 ; APPLICANT: Burton, Frank H.; Sutcliffe, Gregor  
 ; TITLE OF INVENTION: CHOLERA TOXIN GENE REGULATED BY GROWTH  
 ; HORMONE PROMOTER  
 ; NUMBER OF SEQUENCES: 18  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/528,852  
 ; FILING DATE: 18-MAY-1990  
 ; SEQ ID NO: 3  
 ; LENGTH: 124  
 5223610-3

Query Match 88.7%; Score 94; DB 6; Length 124;  
Best Local Similarity 90.5%; Pred. NO. 3.3e-09;  
Matches 19; Conservative 0; Mismatches 2; Indels

Qy	1	GATFQVEVPGSQHIDSQKAI	21
Dp	66	GAIFOVEVPSSOHIDSQKAI	86

```

RESULT 21
US-08-878-989-2
: Sequence 2, Application US/08878989
: Patent No. 5885803
: GENERAL INFORMATION:
: APPLICANT: Bandman, Olga
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Corley, Neil C.
: APPLICANT: Guegler, Karl G.
: APPLICANT: Lal, Preeti
: APPLICANT: Goli, Surva K.
: APPLICANT: Shah, Purvi
: TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN
: TITLE OF INVENTION: KINASES
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/878,989
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J J
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0321 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: TELEX:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 448 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: TBLYN0701

```

QY 1 GATFQVEVPGSQHIDSQKKA 20  
| : | | | : | | |

CLASSIFICATION: 1  
 PRIOR APPLICATION DATA: 2  
 APPLICATION NUMBER: 08/480,351 3  
 FILING DATE: 4  
 ATTORNEY/AGENT INFORMATION: 5  
 NAME: Kagan, Sarah A. 6  
 REGISTRATION NUMBER: 32,141 7  
 REFERENCE/DOCKET NUMBER: 1107,57434 8  
 TELECOMMUNICATION INFORMATION: 9  
 TELEPHONE: 202-508-9100 10  
 TELEFAX: 202-508-9299 11  
 INFORMATION FOR SEQ ID NO: 2: 12  
 SEQUENCE CHARACTERISTICS: 13

;  
; LENGTH: 856 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; US-08-709-784-2

Query Match 38.2%; Score 40.5; DB 3; Length 856;  
Best Local Similarity 75.0%; Pred. No. 85;  
Matches 9; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

QY 10 GSO-HIDSQKKA 20  
||| |::||:|  
Db 528 GSQEHVDSQKKA 539

RESULT 25  
US-08-209-521-23  
; Sequence 23, Application US/08209521  
; Patent No. 5922855  
; GENERAL INFORMATION:  
; APPLICANT: Liskay, Robert M.  
; APPLICANT: Bronner, C. Eric  
; APPLICANT: Baker, Sean M.  
; APPLICANT: Bollag, Roni J.  
; APPLICANT: Kolodner, Richard D.  
; TITLE OF INVENTION: MAMMALIAN DNA MISMATCH REPAIR GENES  
; TITLE OF INVENTION: hMLH1 AND hPMS1  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kolisch, Hartwell, Dickinson, McCormack &  
; ADDRESSEE: Heuser  
; STREET: 520 S.W. Yamhill, Suite 200  
; CITY: Portland  
; STATE: Oregon  
; COUNTRY: US  
; ZIP: 97204  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/209,521  
; FILING DATE: 08-MAR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Van Rysseberghe, Pierre C.  
; REGISTRATION NUMBER: 33,557  
; REFERENCE/DOCKET NUMBER: OHSU 306A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (503) 224-6655  
; TELEFAX: (503) 295-6679  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 862 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-209-521-23

Query Match 38.2%; Score 40.5; DB 2; Length 862;  
Best Local Similarity 75.0%; Pred. No. 86;  
Matches 9; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

QY 10 GSO-HIDSQKKA 20

Db 528 GSQEHVDSQKKA 539  
||| |::||:|

Search completed: July 16, 2001, 16:36:18  
Job time: 207 sec



***This Page Blank (uspto)***

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 16:37:01 ; Search time 34.9 Seconds  
(without alignments)  
45.836 Million cell updates/sec

Title: US-09-786-648-4  
Perfect score: 106  
Sequence: 1 GATQVEVPGSHDSOKKAI 21  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR58:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	95.3	124	1 QLECB	heat-labile entero
2	100	94.3	124	1 XVVCB	cholera enterotoxi
3	48	45.3	392	2 T04150	RAD23 protein homo
4	46	43.4	374	2 T19866	hypothetical prote
5	46	43.4	574	2 C86400	hypothetical prote
6	44.5	42.0	374	2 D81715	conserved hypothet
7	44	41.5	91	2 C96580	hypothetical prote
8	44	41.5	255	2 A86457	hypothetical prote
9	44	41.5	1742	2 S76110	hypothetical prote
10	43	40.6	365	2 H69231	sensory transducti
11	43	40.6	399	2 T46898	queuine tRNA-ribos
12	43	40.6	439	2 T45189	kinesin heavy chai
13	42.5	40.1	376	1 S17246	chorismate synthas
14	42	39.6	216	2 F83962	hypothetical prote
15	42	39.6	299	2 JG0178	chitinase (EC 3.2.
16	42	39.6	427	2 JC5694	stress-activated p
17	42	39.6	787	2 G81692	inner membrane pro
18	42	39.6	864	2 T49574	probable carnitine
19	41.5	39.2	500	2 JC4022	4-aminobutyrate tr
20	41.5	39.2	1090	2 S11823	alpha-dextrin endo
21	41	38.7	91	2 F81109	conserved hypothet
22	41	38.7	128	1 U0D0R	ubiquitin / riboso
23	41	38.7	154	1 U0D0R7	ubiquitin / riboso
24	41	38.7	228	2 D34080	ubiquitin 18 - sil
25	41	38.7	229	2 B27806	ubiquitin (clone 1
26	41	38.7	380	2 C34080	polyubiquitin 5 (c
27	41	38.7	380	2 C34080	polyubiquitin 5 (c
28	41	38.7	381	2 A27806	polyubiquitin 5 (c
29	41	38.7	386	1 S66056	yaan protein - Bac

30	41	38.7	532	2 A34080	polyubiquitin 7 (c
31	41	38.7	550	2 T01770	hypothetical prote
32	41	38.7	2911	2 T20566	hypothetical prote
33	40.5	38.2	862	2 S47598	mutL protein homol
34	40	37.7	105	2 G69903	hypothetical prote
35	40	37.7	227	2 T32894	hypothetical prote
36	40	37.7	242	2 T34767	hypothetical prote
37	40	37.7	260	2 C96827	protein F20B17.2 (
38	40	37.7	355	2 T24938	hypothetical prote
39	40	37.7	367	2 T01751	gibberellin 20-oxi
40	40	37.7	368	2 T04861	hypothetical prote
41	40	37.7	378	2 T52658	thiosulfate sulfur
42	40	37.7	379	2 T01034	thiosulfate sulfur
43	40	37.7	461	2 T16225	hypothetical prote
44	40	37.7	603	2 T16655	hypothetical prote
45	40	37.7	655	2 T39064	RNA binding protei

ALIGNMENTS

RESULT 1

QLECB

heat-labile enterotoxin chain B precursor - Escherichia coli

C:Species: Escherichia coli

C>Date: 29-Jun-1981 #sequence\_revision 29-Jun-1981 #text\_change 18-Jun-1999

C:Accession: A01820; B26946; I41194; I41287; I67644; A61475

R:Dallas, W.S.; Falkow, S.

Nature 288, 499-501, 1980

A:Title: Amino acid sequence homology between cholera toxin and Escherichia coli heat

A:Reference number: A01820; MUID:81074965

A:Accession: A01820

A:Molecule type: mRNA

A:Residues: 1-124 <DAL>

R:Yamamoto, T.; Gojobori, T.; Yokota, T.

J. Bacteriol. 169, 1352-1357, 1987

A:Title: Evolutionary origin of pathogenic determinants in enterotoxigenic Escherich

A:Reference number: A26946; MUID:87137303

A:Accession: B26946

A:Molecule type: DNA

A:Residues: 1-27, 'E', '29-63', 'K', '65-124 <YAM>

A:Cross-references: EMBL:M15363; NID:g148335; PIDN:AAA24792.1; PID:g148336

R:Leong, J.; Vinal, A.C.; Dallas, W.S.

Infect. Immun. 48, 73-77, 1985

A:Title: Nucleotide sequence comparison between heat-labile toxin B-subunit cistrons

A:Reference number: I41194; MUID:85156481

A:Accession: I41194

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-5, 'F', '7-17', 'C', '19-24', 'S', '26-27', 'E', '29-33', 'H', '35-63', 'K', '65-66', 'A', '68-122

A:Cross-references: GB:M17874; NID:g145830; PIDN:AAA98064.1; PID:g145831

R:Ibrahim, I.; Gentz, R.

J. Biol. Chem. 262, 10189-10194, 1987

A:Title: A functional interaction between the signal peptide and the translation appa

tulum.

A:Reference number: I41287; MUID:87280041

A:Accession: I41287

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-22 <RE2>

A:Cross-references: GB:M17101; NID:g146375; PIDN:AAA23973.1; PID:g146376

R:Inoue, T.; Tsuji, T.; Koto, M.; Imamura, S.; Miyama, A.

FEMS Microbiol. Lett. 108, 157-161, 1993

A:Title: Amino acid sequence of heat-labile enterotoxin from chicken enterotoxigenic

A:Reference number: I53542; MUID:93252225

A:Accession: I67644

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-17, 'C', '19', 'Y', '21-24', 'S', '26-27', 'E', '29-63', 'K', '65-66', 'A', '68-122', 'E', '124 <R

A:Cross-references: GB:S60731; NID:g408994; PIDN:AAC60441.1; PID:g408996

R:Tsuji, T.; Lida, T.; Honda, T.; Miwatani, T.; Nagahama, M.; Sakurai, J.; Wada, K.;

Microb. Pathog. 2, 381-390, 1987

A:Title: A unique amino acid sequence of the B subunit of a heat-labile enterotoxin isolated from *Vibrio cholerae* O139  
A:Reference number: A61475; MUID:89180953  
A:Accession: A61475  
A:Molecule type: protein  
A:Residues: 22-24,'S',26-27,'E',29-63,'K',65-66,'A',68-95,'A',97-122,'E',124 <TSU>  
A:Experimental source: Strain 240-3  
C:Function: the heat-labile enterotoxin molecule contains one A chain and five or six B chains  
A:Description: the biological activity of the toxin is produced by the A chain, which acts as a superfamily: cholera enterotoxin beta chain  
C:Keywords: enterotoxin  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-124/Product: heat-labile enterotoxin chain B #status predicted <MAT>  
F:30-107/Disulfide bonds: #status predicted

Query Match 95.3%; Score 101; DB 1; Length 124;  
Best Local Similarity 95.2%; Pred. No. 1.6e-09;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSQKKAI 21  
DB 56 GATFOVEVPGSQHIDSQKKAI 86

RESULT 2  
XVVCB  
cholera enterotoxin, B chain precursor VCL456 [validated] - Vibrio cholerae (strain N169)  
A:Alternate names: enterotoxin beta chain  
C:Species: Vibrio cholerae  
C:Date: 24-Apr-1984 #sequence\_revision 01-Sep-2000 #text\_change 02-Feb-2001  
C:Accession: S14624; S39238; S39241; H82196; JCI1078; S17666; PC1010; A05130; A01819; A381819  
R:Dams, E.; de Wolf, M.; Dierick, W.  
submitted to the EMBL Data Library, March 1991  
A:Description: Correction of the cholera toxin nucleotide sequence of the Vibrio cholerae  
A:Reference number: S14623  
A:Accession: S14624  
A:Molecule type: DNA  
A:Residues: 1-124 <DAM>  
A:Cross-references: EMBL:X58786; NID:g48420; PIDN:CAA1593.1; PID:g48422  
A:Experimental source: strain 2125  
R:Lebens, M.; Holmgren, J.  
submitted to the EMBL Data Library, November 1993  
A:Description: Structure and arrangement of the Cholera toxin genes in vibrio Cholera O1  
A:Reference number: S39238  
A:Accession: S39238  
A:Molecule type: DNA  
A:Residues: 1-124 <LEB>  
A:Cross-references: EMBL:X76390; NID:g433856; PIDN:CAA53973.1; PID:g433857  
A:Accession: S39241  
A:Molecule type: DNA  
A:Residues: 1-124 <LEW>  
A:Cross-references: EMBL:X76391; NID:g433859; PIDN:CAA53976.1; PID:g433861  
R:Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, H.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833  
A:Accession: H82196  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-124 <HEI>  
A:Cross-references: GB:AE003852; NID:g9655952; PIDN:AAF94613.1; GSPDB:GN001  
A:Experimental source: serogroup O1; strain N15961; biotype El Tor  
R:Shi, C.H.; Cao, C.; Zhang, J.S.; Ma, Q.J.  
Chinese Biochem. J. 9, 395-399, 1993  
A:Title: Nucleotide sequence analysis of the gene encoding the classical biotype V. cholerae  
A:Reference number: JCI1078  
A:Accession: JCI1078  
A:Molecule type: DNA  
A:Residues: 1-20,'Q',22-31,'Q',33-38,'H',40-49,'G',51-67,'T',69-124 <SHI>  
A:Experimental source: classical biotype strain 569B

R:Dams, E.; de Wolf, M.; Dierick, W.  
Biochim. Biophys. Acta 1090, 139-141, 1991  
A:Title: Nucleotide sequence analysis of the CT operon of the Vibrio cholerae classic  
A:Reference number: S17665; MUID:91355224  
A:Accession: S17666  
A:Molecule type: DNA  
A:Residues: 1-38,'H',40-67,'T',69-124 <DA2>  
A:Cross-references: EMBL:X58785; NID:g48888; PIDN:CAA1591.1; PID:g48890  
R:Ma, Q.J.; Liu, C.X.; Xiong, L.S.; Yu, X.Q.  
Sci. Sin. B Chem. Biol. Agric. Med. Earth Sci. 34, 274-280, 1991  
A:Title: B subunit of cholera toxin produced in Escherichia coli.  
A:Reference number: PC1010  
A:Accession: PC1010  
A:Molecule type: protein  
A:Residues: 22-38,'H',40-41 <MAQ>  
R:Mekalanos, J.J.; Swartz, D.J.; Pearson, G.D.N.; Harford, N.; Groyne, F.; de Wilde, N.  
Nature 306, 551-557, 1983  
A:Reference number: A93320; MUID:84068199  
A:Accession: A05130  
A:Molecule type: DNA  
A:Residues: 1-32,'S',34-74,'S',76-124 <MEK>  
A:Cross-references: GB:X00171; NID:g48347; PIDN:CAA24996.1; PID:g758351  
R:Kurosky, A.; Markel, D.E.; Peterson, J.W.  
J. Biol. Chem. 252, 7257-7264, 1977  
A:Title: Covalent structure of the beta chain of cholera enterotoxin.  
A:Reference number: A01819; MUID:78005537  
A:Accession: A01819  
A:Molecule type: protein  
A:Residues: 22-38,'H',40-42,'N',44-67,'T',69-90,'N',92-124 <KUR>  
R:Lai, C.Y.  
J. Biol. Chem. 252, 7249-7256, 1977  
A:Title: Determination of the primary structure of cholera toxin B subunit.  
A:Reference number: A38033; MUID:78005536  
A:Accession: A38033  
A:Molecule type: protein  
A:Residues: 22-38,'H',40-42,'N',44-67,'T',69,'E',71-90,'N',92-124 <IAI>  
A:Note: the difference at residue 70 may be due to deamidation during preparation  
R:Nakashima, Y.; Napiorkowski, P.; Schafer, D.E.; Konigsberg, W.H.  
FEBS Lett. 69, 275-278, 1976  
A:Title: Primary structure of the B subunit of cholera enterotoxin.  
A:Reference number: A38034; MUID:77026365  
A:Accession: A38034  
A:Molecule type: protein  
A:Residues: 22-38,'H',40-67,'T',69,'E',71,'QS',74-75,'VE',78-86,'Q',88-99,'Q',101-103  
R:Takao, T.; Watanabe, H.; Shimonishi, Y.  
Eur. J. Biochem. 146, 503-508, 1985  
A:Title: Facile identification of protein sequences by mass spectrometry.  
A:Reference number: A21910; MUID:85126976  
A:Accession: A21910  
A:Molecule type: protein  
A:Residues: 22-38,'H',40-42,'N',44-67,'T',69-90,'N',92-124 <YAK>  
A:Experimental source: biotype Inaba 569B  
A:Note: Asn-65 was partially deaminated to Asp  
C:Comment: The authors translated the codon TCA for residue 33 as Tyr.  
C:Genetics:  
A:Gene: VC1456  
A:Map position: 1  
C:Complex: the cholera enterotoxin molecule contains three kinds of chains; an alpha chain noncovalently with the subunit B, an aggregate of five beta chains  
C:Function:  
A:Description: involved in binding of the toxin to cell membranes  
C:Superfamily: cholera enterotoxin beta chain  
C:Keywords: enterotoxin; toxin  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-124/Product: cholera enterotoxin chain B #status predicted <MAT>  
F:30-107/Disulfide bonds: #status experimental

Query Match 94.3%; Score 100; DB 1; Length 124;  
Best Local Similarity 95.2%; Pred. No. 2.4e-09;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSQKKAI 21



[illegible]

A:Reference number: A86141; MUID:21016719  
A:Accession: C96580  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-91 <STO>  
A:Cross-references: GB:AE005173; NID:g4587538; PIDN:AAD25769.1; GSPDB:GN00141  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001  
C:Genetics:  
A:Gene: F1511.6  
A:Map position: 1

Query Match 41.5%; Score 44; DB 2; Length 91;  
Best Local Similarity 47.6%; Pred. No. 3.6;  
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSQKAI 21  
| | | : | | | : |  
DB 58 GRTFNLEVGSEIIQQVKMI 78

RESULT 8  
A86457  
hypothetical protein AAG21605.1 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001  
C:Accession: A86457  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: A86457  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-255 <STO>  
A:Cross-references: GB:AE005172; NID:gl0645493; PIDN:AAG21605.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 41.5%; Score 44; DB 2; Length 255;  
Best Local Similarity 42.1%; Pred. No. 11;  
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 3 TFOVEVPGSQHIDSQKAI 21  
| | | : | | | : |  
DB 102 TFRVSGPGQHRNKRDSAV 120

RESULT 9  
S76110  
hypothetical protein - Synecocystis sp. (strain PCC 6803)  
C:Species: Synecocystis sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
C:Accession: S76110  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis  
s.  
A:Reference number: S74322; MUID:97061201  
A:Accession: S76110  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1742 <KAN>  
A:Cross-references: EMBL:D63999; GB:AB001339; NID:gl001396; PIDN:BAAL0088.1; PID:d101073

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Genetics:  
A:Start codon: GTG

Query Match 41.5%; Score 44; DB 2; Length 1742;  
Best Local Similarity 38.1%; Pred. No. 87;  
Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSQKAI 21  
| | | : | | | : |  
DB 121 GGTFSIDVPGSALVADPORTI 141

RESULT 10  
H69231  
sensory transduction histidine kinase - Methanobacterium thermoautotrophicum (strain  
C:Species: Methanobacterium thermoautotrophicum  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
C:Accession: H69231  
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.  
; Old, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwan,  
K.; S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu  
A:Reference number: A69000; MUID:98037514  
A:Accession: H69231  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-365 <MTH>  
A:Cross-references: GB:AE000872; GB:AE000666; NID:g2622082; PIDN:AAB85482.1; PID:g2622  
A:Experimental source: strain Delta H  
C:Genetics:  
A:Gene: MTH985

Query Match 40.6%; Score 43; DB 2; Length 365;  
Best Local Similarity 36.8%; Pred. No. 24;  
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 2 ATFQVEVPGSQHIDSQKKA 20  
| | | : | | | : |  
DB 49 AIFLVKIPGGLVDANRSA 67

RESULT 11  
T46898  
queuine tRNA-ribosyltransferase (EC 2.4.2.29) [validated] - Zymomonas mobilis  
N:Alternate names: tRNA guanine transglycosylase  
C:Species: Zymomonas mobilis  
C:Date: 17-Mar-2000 #sequence\_revision 17-Mar-2000 #text\_change 18-Aug-2000  
C:Accession: T46898; T46899  
R:Reuter, K.; Riemer, R.  
J. Bacteriol. 177, 5284-5288, 1995  
A:Title: Sequence analysis and overexpression of the Zymomonas mobilis tgt gene encod  
A:Reference number: 224129; MUID:95394847  
A:Accession: T46898  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-399 <REU>  
A:Cross-references: EMBL:L33777; PIDN:AAA27704.1  
A:Accession: T46899  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 5-399 <RE2>  
A:Cross-references: EMBL:L33777; PIDN:AAA27705.1  
C:Genetics:  
A:Gene: tgt  
C:Complex: monomer [validated; MUID:95394847]  
C:Function:  
A:Description: (EC 2.4.2.29) [validated; MUID:95394847]; catalyzes the exchange of gu  
ine  
A:Pathway: queuosine biosynthesis; tRNA modification

F:347-363/Domain: transmembrane #status predicted <TMM>

Query Match 40.1%; Score 42.5; DB 1; Length 376;  
Best Local Similarity 62.5%; Pred. No. 29;  
Matches 10; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

Qy	1	GATFQ-VEVPGSQHID	15
		↑: ↑↑↑↑↑↑↑↑↑↑	
Db	258	GSGFQGVSVPGSKHND	273

RESULT 14

FB3962

hypothetical protein BH2502 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 31-Dec-2000

C:Accession: F83962

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: A83650; MUID:20263314

A:Accession: F83962

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-216 <STO>

A:Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BA06221.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

C:Gene: BH2502

C:Superfamily: yeast ribulose-5-phosphate-epimerase

Query Match	39.6%	Score 42;	DB 2;	Length 216;
Best Local	Similarity 33.3%;	Pred. No. 20;		
Matches	7;	Conservative	7;	Mismatches
			7;	Indels
				Gaps
				0

QY 1 GATFQVEVPGSQHIDSQKAI 21  
| | | | | : : : : :  
Db 166 GLTFEIEVDGGVNEETAKQCV 186

```

RESULT 15
JG0178
chitinase (EC 3.2.1.14) Chib1 - soybean
C:Species: Glycine max (soybean)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 21-Jul-2000
C:Accession: JG0178
R:Watanabe, A.; Nong, V.H.; Zhang, D.; Arahira, M.; Yeboah, N.A.; Udaoka, K.; Fujisaki, Y.
Biosci. Biotechnol. Biochem. 63, 251-256, 1999
A:Title: Molecular cloning and ethylene-inducible expression of Chib1 chitinase
A:Reference number: JG0178; MUID:99208985
A:Accession: JG0178
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-299 <NAT>
A:Cross-references: DDBJ:AB006748; NID:g4835581; PIDN:BA47675.1; PID:g4835582
C:Superfamily: plant chitinase III
C:Keywords: glycosidase; hydrolase

Query Match 39.6%; Score 42; DB 2; Length 299;
Best Local Similarity 38.1%; Pred. No. 28;
Matches 8; Conservative 3; Mismatches 10; Indels 0; Gaps 0

QY 1 GATFQVEVPGVSGHIDSQKKAI 21
      | | | | | | | |
Db 147 GIDFDIESGGSDHYDDLARAL 167

RESULT 16
JC5694
stress-activated protein kinase (EC 2.7.1.-) JNKB - common carp

```

A:Experimental source: BAC clone B208; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B208.220  
A:Map position: 6  
A:Introns: 138/1

Query Match 39.6%; Score 42; DB 2; Length 864;  
Best Local Similarity 44.4%; Pred. No. 88;  
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 3 TFOVEVPGSQHIDSQKKA 20  
|| : || || : ||  
DB 542 TFWADFCDAKIDALRKA 559

RESULT 19  
JC4022  
4-aminobutyrate transaminase (EC 2.6.1.19) - human  
N:Alternate names: beta-alanine--oxoglutarate transaminase  
C:Species: Homo sapiens (man)  
C:Date: 13-Jun-1995 #sequence\_revision 14-Jul-1995 #text\_change 18-Jun-1999  
C:Accession: JC4022  
R:Oselt, I.D.; Churchich, J.E.  
Gene 155, 185-187, 1995  
A:Title: Screening and sequence determination of a cDNA encoding the human brain 4-amino-  
A:Reference number: JC4022; MUID:95237607  
A:Accession: JC4022  
A:Molecule type: mRNA  
A:Residues: 1-500 <OE>  
A:Cross-references: GB:I32961; NID:g602704; PIDN:AAA74449.1; PID:g602705  
A:Experimental source: brain  
C:Comment: This enzyme catalyzes the conversion of 4-aminobutyrate, the inhibitory ne  
C:Genetics:  
A:Gene: GDB:ABAT; GABAT  
A:Cross-references: GDB:581658; OMIM:137150  
A:Map position: 19q13-19q13  
C:Superfamily: 4-aminobutyrate transaminase  
C:Keywords: aminotransferase

Query Match 39.2%; Score 41.5; DB 2; Length 500;  
Best Local Similarity 55.6%; Pred. No. 59;  
Matches 10; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 3 TFOVEVPGSQHIDSQKKA 20  
|::: ||||:| || |  
DB 17 TYRLVPGSRHI-SQAAA 33

RESULT 20  
S11823  
alpha-dextrin endo-1,6-alpha-glucosidase (EC 3.2.1.41) precursor - Klebsiella pneumonia  
N:Alternate names: pullulanase  
C:Species: Klebsiella pneumoniae  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 08-Oct-1999  
C:Accession: S11823; A32880  
R:Kornacker, M.G.; Pugsley, A.P.  
Mol. Microbiol. 4, 73-85, 1990  
A:Title: Molecular characterization of pula and its product, pullulanase, a secreted  
A:Reference number: S11823; MUID:90205629  
A:Accession: S11823  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1090 <KOR>  
A:Cross-references: EMBL:X52181; NID:g43912; PIDN:CAA36431.1; PID:g43913  
R:D'Enfert, C.; Pugsley, A.P.  
J. Bacteriol. 171, 3673-3679, 1989  
A:Title: Klebsiella pneumoniae pulS gene encodes an outer membrane lipoprotein requir  
A:Reference number: A32880; MUID:89291709  
A:Accession: A32880  
A>Status: preliminary  
A:Molecule type: DNA

A:Residues: 944-1090 <DN>  
A:Cross-references: GB:M29097; NID:g342046; PIDN:AAA61976.1; PID:g643577  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 39.2%; Score 41.5; DB 2; Length 1090;  
Best Local Similarity 38.5%; Pred. No. 1.4e+02;  
Matches 10; Conservative 4; Mismatches 7; Indels 5; Gaps 1;

QY 1 GATFQVEVPGSQHID-----SOKKAI 21  
||| ||| : |||  
DB 312 GVTRVWAPTAQQVDVVVYSADKKVI 337

RESULT 21  
F81109  
conserved hypothetical protein NMB1218 [imported] - Neisseria meningitidis (strain MC58)  
C:Species: Neisseria meningitidis  
C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 02-Feb-2001  
C:Accession: F81109; F81907  
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: AB1000; MUID:20175755  
A:Accession: F81109  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-91 <ET>  
A:Cross-references: GB:AE002469; GB:AE002098; NID:g7226446; PIDN:AAF1600.1; PID:g722645  
A:Experimental source: serogroup B, strain MC58  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
Holtroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: AB1775; MUID:20222556  
A:Accession: F81907  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-91 <PAR>  
A:Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84626.1; PID:g738004  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMB1218; NMA1380

Query Match 38.7%; Score 41; DB 2; Length 91;  
Best Local Similarity 38.1%; Pred. No. 11;  
Matches 8; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 1 GATFQVEVPGSQHIDSQKKAI 21  
||| ||| : |||  
DB 61 GATVQVKVENQGLDNIYRAL 81

RESULT 22  
U0DOR  
ubiquitin / ribosomal protein CEP52 - slime mold (Dictyostellium discoideum)  
N:Alternate names: ubiquitin fusion protein  
C:Species: dictyostellium discoideum  
C>Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 22-Jun-1999  
C:Accession: S00357; A25863  
R:Mueller-Taubenberger, A.; Westphal, M.; Jaeger, E.; Noegel, A.; Gerisch, G.  
FEBS Lett. 229, 273-278, 1988  
A:Title: Complete cDNA sequence of a Dictyostellium ubiquitin with a carboxy-terminal tail  
A:Reference number: S00357; MUID:88152253  
A:Accession: S00357  
A:Molecule type: mRNA  
A:Residues: 1-128 <ME>  
A:Cross-references: EMBL:X07210; NID:g7381; PIDN:CAA30183.1; PID:g7382  
A:Experimental source: strain AX2-214

C:Genetics:  
A:Gene: DUB1  
C:Superfamily: ubiquitin/ribosomal protein CEP52; ribosomal protein CEP52 homology; u  
C:Keywords: DNA binding; protein biosynthesis; protein degradation; ribosome; zinc fi  
F:1-76/Product: ubiquitin #status predicted <UBI>  
F:1-76/Domain: ubiquitin homology <UBH>  
F:77-128/Product: ribosomal protein CEP52 #status predicted <RIB>  
F:77-128/Domain: ribosomal protein CEP52 homology <CPH>  
F:95-114/Region: zinc finger CCCC motif  
F:121-128/Region: nuclear location signal

Query Match 38.7%; Score 41; DB 1; Length 128;  
Best Local Similarity 42.9%; Pred. No. 16;  
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 GATFQVEVPGSQHIDSQKKAI 21  
||| ||| : |||  
DB 10 GXTITLEVEGSDNIENVKAKI 30

RESULT 23  
U0DOR7  
ubiquitin / ribosomal protein S27a - slime mold (Dictyostellium discoideum)  
C:Species: dictyostellium discoideum  
C>Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 21-Jul-2000  
C:Accession: E34080  
R:Ohmachi, T.; Giorda, R.; Shaw, D.R.; Ennis, H.L.  
Biochemistry 28, 5226-5231, 1989  
A:Title: Molecular organization of developmentally regulated Dictyostellium discoideum  
A:Reference number: A34080; MUID:89352609  
A:Accession: E34080  
A:Molecule type: mRNA  
A:Residues: 1-154 <OHM>  
A:Cross-references: GB:M23750; GB:J02858; NID:gl67940; PIDN:AAA33264.1; PID:gl67941  
C:Superfamily: ubiquitin / rat ribosomal protein S27a; ribosomal protein S27a homology  
C:Keywords: protein biosynthesis; protein degradation; ribosome  
F:1-76/Product: ubiquitin #status predicted <NAT1>  
F:1-76/Domain: ubiquitin homology <UBH>  
F:77-154/Product: ribosomal protein S27a #status predicted <NAT2>  
F:102-152/Domain: ribosomal protein S27a homology <RIB>

Query Match 38.7%; Score 41; DB 1; Length 154;  
Best Local Similarity 42.9%; Pred. No. 20;  
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 GATFQVEVPGSQHIDSQKKAI 21  
||| ||| : |||  
DB 10 GXTITLEVEGSDNIENVKAKI 30

RESULT 24  
D34080  
ubiquitin 18 - slime mold (Dictyostellium discoideum)  
C:Species: dictyostellium discoideum  
C>Date: 30-Mar-1990 #sequence\_revision 30-Mar-1990 #text\_change 23-Feb-1997  
C:Accession: D34080  
R:Ohmachi, T.; Giorda, R.; Shaw, D.R.; Ennis, H.L.  
Biochemistry 28, 5226-5231, 1989  
A:Title: Molecular organization of developmentally regulated Dictyostellium discoideum  
A:Reference number: A34080; MUID:89352609  
A:Accession: D34080  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-228 <OHM>  
A:Cross-references: GB:M23751  
C:Superfamily: polyubiquitin 3; ubiquitin homology  
C:Keywords: duplication; nucleus; polyprotein; protein degradation  
F:1-76/Domain: ubiquitin homology <UBH1>  
F:77-152/Domain: ubiquitin homology <UBH2>  
F:153-228/Domain: ubiquitin homology <UBH3>

Query Match 38.7%; Score 41; DB 2; Length 228;  
 Best Local Similarity 42.9%; Pred. No. 31;  
 Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GATFOVEVPGSQHIDSQKKAI 21  
 | | :|| || :|: |  
 Db 10 GKTITLEVEGSDNIENVKAKI 30

RESULT 25

B27806  
 ubiquitin (clone lambda229) - slime mold (Dictyostelium discoideum)  
 C:Species: Dictyostelium discoideum  
 C:Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 06-Feb-1998  
 C:Accession: B27806  
 R:Giorda, R.; Ennis, H.L.  
 Mol. Cell. Biol. 7, 2097-2103, 1987  
 A:Title: Structure of two developmentally regulated Dictyostelium discoideum ubiquitin g  
 A:Reference number: A27806; MUID:87257921  
 A:Accession: B27806  
 A:Molecule type: mRNA  
 A:Residues: 1-229 <GIO>  
 A:Cross-references: GB:M23754  
 C:Superfamily: polyubiquitin 3; ubiquitin homology  
 C:Keywords: duplication; nucleus; polyprotein; protein degradation  
 F:1-76/Domain: ubiquitin homology <UBH1>  
 F:77-152/Domain: ubiquitin homology <UBH2>  
 F:153-228/Domain: ubiquitin homology <UBH3>

Query Match 38.7%; Score 41; DB 2; Length 229;  
 Best Local Similarity 42.9%; Pred. No. 31;  
 Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GATFOVEVPGSQHIDSQKKAI 21  
 | | :|| || :|: |  
 Db 10 GKTITLEVEGSDNIENVKAKI 30

Search completed: July 16, 2001, 16:37:02  
 Job time: 211 sec

GenCore version 4.5

Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 16:44:03 ; Search time 19.88 Seconds  
(without alignments)  
36.185 Million cell updates/sec

Title: US-09-786-648-4

Perfect score: 106

Sequence: 1 GATFOVEVPGSQHDSQKAI 21

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	100.0	124	1	ELBH_ECOLI
2	101	95.3	124	1	ELBP_ECOLI
3	100	94.3	124	1	CHPB_VIBCH
4	44	41.5	467	1	IRP6_HUMAN
5	43	40.6	385	1	TGF_ZYMMO
6	42.5	40.1	376	1	AROC_YEAST
7	42	39.6	855	1	ST14_HUMAN
8	41.5	39.2	500	1	GABT_HUMAN
9	41.5	39.2	1090	1	PULA_KLEPN
10	41	38.7	76	1	UBIQ_DICDI
11	41	38.7	313	1	YP88_CAEEL
12	41	38.7	386	1	YAAN_BACSU
13	40.5	38.2	862	1	PMS2_HUMAN
14	40	37.7	288	1	HS74_CAEEL
15	40	37.7	355	1	YRY1_CAEEL
16	40	37.7	655	1	YDH2_SCHPO
17	40	37.7	683	1	RPC_BPHPC
18	40	37.7	2733	1	RRPB_CVMA5
19	39.5	37.3	560	1	INR1_SHEEP
20	39	36.8	112	1	Y13K_SSV1
21	39	36.8	263	1	YC43_ODOSI
22	39	36.8	456	1	SR54_THEAC
23	39	36.8	459	1	IL7R_MOUSE
24	39	36.8	491	1	CD5_RAT
25	39	36.8	661	1	HS7C_CAEEL
26	39	36.8	684	1	CC4_CANAL
27	39	36.8	704	1	PCCA_RAT
28	39	36.8	775	1	TH11_SCHPO
29	39	36.8	842	1	DPOL_HPBVM
30	39	36.8	2731	1	RRPB_CVMJH
31	38.5	36.3	293	1	CH7A_CICAR
32	38.5	36.3	344	1	LINA_PSES5
33	38.5	36.3	529	1	TACY_LISMO

34 38.5 36.3 744 1 CATA\_EMENI  
35 38 35.8 76 1 UBIQ\_NEUCR  
36 38 35.8 148 1 IF5A\_AERPE  
37 38 35.8 169 1 Y459\_MYCGE  
38 38 35.8 169 1 Y459\_MYCPN  
39 38 35.8 205 1 HAP3\_KLULA  
40 38 35.8 259 1 MOB2\_YEAST  
41 38 35.8 359 1 THD1\_SOLTU  
42 38 35.8 417 1 PGK1\_RHINI  
43 38 35.8 454 1 VNUC\_THOYG  
44 38 35.8 523 1 DBP3\_YEAST  
45 38 35.8 674 1 COAA\_BACTJ

P55305 emericella  
P13117 neurospora  
Q9YA53 aeropyrum p  
Q49436 mycoplasma  
P75118 mycoplasma  
P40914 kluyveromyc  
P43563 saccharomyc  
P31212 solanum tub  
P29405 rhizopus ni  
P89216 thogoto vir  
P20447 saccharomyc  
O87905 bacillus th

## ALIGNMENTS

## RESULT 1

ELBH\_ECOLI STANDARD; PRT; 124 AA.  
AC P13811;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE HEAT-LABILE ENTEROTOXIN B CHAIN PRECURSOR (LT-B, HUMAN) (LTH-B).  
GN ELTB OR LTPB.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ISOLATE H74-114;  
RX MEDLINE=85156481; PubMed=3884513;  
RA Leong J., Vinal A.C., Dallas W.S.;  
RT "Nucleotide sequence comparison between heat-labile toxin B-subunit  
cistrons from Escherichia coli of human and porcine origin.";  
RL Infect. Immun. 48:73-77(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ISOLATE H10407;  
RX MEDLINE=83114628; PubMed=6759877;  
RA Yamamoto T., Tamura T.A., Yokota T.;  
RT "Overlapping genes in the heat-labile enterotoxin operon originating  
from Escherichia coli human strain.";  
RL Mol. Gen. Genet. 188:356-359(1982).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ISOLATE H10407;  
RX MEDLINE=93252225; PubMed=8486242;  
RA Inoue T., Tsuji T., Koto M., Imamura S., Miyama A.;  
RT "Amino acid sequence of heat-labile enterotoxin from chicken  
enterotoxigenic Escherichia coli is identical to that of human strain  
H 10407.";  
RL FEMS Microbiol. Lett. 108:157-161(1993).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ETEC LT 87;  
RA Germani Y., Desperrier J.M.;  
RT Submitted (Oct-1996) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP DISCUSSION OF SEQUENCE.  
RX MEDLINE=95349400; PubMed=7623669;  
RA Domenighini M., Pizzo M., Jobling M.G., Rappuoli R.;  
RT "Identification of errors among database sequence entries and  
comparison of correct amino acid sequences for the heat-labile  
enterotoxins of Escherichia coli and Vibrio cholerae.";  
RL Mol. Microbiol. 15:1165-1167(1995).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (3.04 ANGSTROMS).  
RX MEDLINE=99185101; PubMed=10085117;  
RA Matkovic-Calogovic D., Loregian A., D'Acunto M.R., Battistutta R.,  
Tossi A., Palu G., Zanotti G.;

"Crystal structure of the B subunit of escherichia coli heat-labile enterotoxin carrying peptides with anti-herpes simplex virus type 1 activity.";  
J. Biol. Chem. 274:8764-8769(1999).  
-!- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.  
-!- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M17874; AAA98064.1; -;  
EMBL; J01646; AAB02982.1; -;  
EMBL; S60731; AAC60441.1; -;  
EMBL; X83966; CAA58800.1; -;  
PDB: 1LTR; 23-MAR-99.  
InterPro: IPR001835; -;  
Pfam: PF01376; Enterotoxin\_B; 1.  
PRINTS; PR00772; ENTEROTOXIN.  
Enterotoxin; Signal; 3D-structure.  
FT SIGNAL 1 21  
CHAIN 22 124 HEAT-LABILE ENTEROTOXIN B CHAIN.  
FT DISULFID 30 107  
FT VARIANT 6 6 F -> C (IN ISOLATE H10407).  
FT VARIANT 20 20 H -> Y (IN ISOLATE H10407).  
FT VARIANT 34 34 H -> R (IN ISOLATE H10407).  
SQ SEQUENCE 124 AA; 14027 MW; E9F7F7C7B9D3BC47 CRC64;  
  
Query Match 100.0%; Score 106; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 2e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GATFQVEVPGSGHDSOKKAI 21  
DB 66 GATFQVEVPGSGHDSOKKAI 86  
  
RESULT 2  
ELBP\_ECOLI STANDARD; PRT; 124 AA.  
AC P32890; P13768; P01557;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE HEAT-LABILE ENTEROTOXIN B CHAIN PRECURSOR (LT-B, PORCINE) (LTP-B).  
GN ELTB OR LTPB.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ISOLATE P307;  
RX MEDLINE=81074965; PubMed=7003397;  
RA Dallas W.S., Falkow S.;  
RT "Amino acid sequence homology between cholera toxin and Escherichia coli heat-labile toxin.";  
RL Nature 288:499-501(1980).  
RN [2]  
RP REVISIONS TO 28 AND 64.  
RC STRAIN-ISOLATE P307;  
RX MEDLINE=85156481; PubMed=3884513;  
RA Leong J., Vinal A.C., Dallas W.S.;  
RT "Nucleotide sequence comparison between heat-labile toxin B-subunit cistrons from Escherichia coli of human and porcine origin.";  
RL Infect. Immun. 48:73-77(1985).  
RN [3]

RP SEQUENCE FROM N.A.  
RC STRAIN-ISOLATE P307;  
RX MEDLINE=87137303; PubMed=3546273;  
RA Yamamoto T., Gotohori T., Yokota T.;  
RT "Evolutionary origin of pathogenic determinants in enterotoxigenic Escherichia coli and Vibrio cholerae O1.";  
RL J. Bacteriol. 169:1352-1357(1987).  
RN [4]  
RP SEQUENCE OF 1-22 FROM N.A.  
RX MEDLINE=87280041; PubMed=3301830;  
RA Ibrahim I., Gentz R.;  
RT "A functional interaction between the signal peptide and the translation apparatus is detected by the use of a single point mutation which blocks translocation across mammalian endoplasmic reticulum";  
RL J. Biol. Chem. 262:10189-10194(1987).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).  
RX MEDLINE=93240541; PubMed=8478941;  
RA Sixma T.K., van Zanten B.A.M., Dauter Z., Hol W.G.J.;  
RT "Refined structure of Escherichia coli heat-labile enterotoxin, a close relative of cholera toxin.";  
RL J. Mol. Biol. 230:890-918(1993).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
RX MEDLINE=91238966; PubMed=2034287;  
RA Sixma T.K., Pronk S.E., Kalk K.H., Wartna E.S., van Zanten B.A.M., Witholt B., Hol W.G.J.;  
RT "Crystal structure of a cholera toxin-related heat-labile enterotoxin from E. coli.";  
RL Nature 351:371-377(1991).  
RN [7]  
RP DISCUSSION OF SEQUENCE.  
RX MEDLINE=95349400; PubMed=7623669;  
RA Domenighini M., Pizzo M., Jobling M.G., Holmes R.K., Rappuoli R.;  
RT "Identification of errors among database sequence entries and comparison of correct amino acid sequences for the heat-labile enterotoxins of Escherichia coli and Vibrio cholerae.";  
RL Mol. Microbiol. 15:1165-1167(1995).  
CC -!- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.  
CC -!- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M17873; AAA98065.1; -;  
EMBL; M15363; AAA24792.1; -;  
EMBL; M17101; AAA23973.1; -;  
PIR; A01820; QLECEB.  
PIR; B26946; QLECEB.  
PDB: 1LTA; 31-JAN-94.  
PDB: 1LTB; 31-JAN-94.  
PDB: 1LTG; 15-SEP-95.  
PDB: 1LTI; 17-AUG-96.  
PDB: 1LTS; 31-JAN-94.  
PDB: 1LTT; 31-JAN-94.  
PDB: 1LTT; 07-JUL-97.  
PDB: 1LT4; 16-JUN-97.  
PDB: 1LT5; 03-DEC-97.  
PDB: 1LT6; 03-DEC-97.  
PDB: 1LTL; 20-APR-95.  
PDB: InterPro; IPR001835; -;  
Pfam: PF01376; Enterotoxin\_B; 1.  
PRINTS; PR00772; ENTEROTOXIN.  
KW Enterotoxin; Signal; 3D-structure.  
FT SIGNAL 1 21  
FT CHAIN 22 124 HEAT-LABILE ENTEROTOXIN B CHAIN.



```

FT DISULFID 30 107
FT TURN 26 30
FT HELIX 31 32
FT STRAND 36 43
FT TURN 47 51
FT STRAND 54 55
FT TURN 58 62
FT TURN 64 65
FT STRAND 68 71
FT TURN 76 77
FT HELIX 80 98
FT TURN 99 100
FT STRAND 103 109
FT STRAND 115 123
SQ SEQUENCE 124 AA; 14133 MW; 6DB7DE58395EA70D CRC64;

Query Match 95.3%; Score 101; DB 1; Length 124;
Best Local Similarity 95.2%; Pred. No. 1.3e-09;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATFQVEVPGSQHIDSQKRAI 21
DB 66 GATFQVEVPGSQHIDSQKRAI 86

RESULT 3
ID CHTB_VIBCH STANDARD; PRT; 124 AA.
AC P01556; Q9JQ02;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CHOLERA ENTEROTOXIN, BETA CHAIN PRECURSOR.
GN CTXB OR TOXB OR VC1456.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=84061784; PubMed=6315707;
RA Lockman H., Kaper J.B.;
RT "Nucleotide sequence analysis of the A2 and B subunits of Vibrio cholerae enterotoxin."
RL J. Biol. Chem. 258:13722-13726(1983).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=EL TOR 2125;
RX MEDLINE=84068199; PubMed=6646234;
RA Mekalanos J.J., Swartz D.J., Pearson G.D.N., Harford N., Groyne F., de Wilde M.;
RT "Cholera toxin genes: nucleotide sequence, deletion analysis and vaccine development."
RL Nature 306:551-557(1983).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=EL TOR 2125;
RA Dams E., de Wolf M., Dierick W.;
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE FROM N.A.
RC STRAIN=4260B / SEROTYPE O139;
RX MEDLINE=94237453; PubMed=8181723;
RA Lebens M., Holmgren J.;
RT "Structure and arrangement of the cholera toxin genes in Vibrio cholerae O139."
RL FEMS Microbiol. Lett. 117:197-202(1994).
RN [5]
RN SEQUENCE FROM N.A.
RC STRAIN=1854 / O139-BENGAL;
RA Yamamoto K., Do V.G.R.F., Xu M., Iida T., Miwatani T., Albert M.J., Honda T.;
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.

[6]
RN SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L., Heideberg J.F., Hickey E.K., Peterson J.D., Unayam L.A., Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D., Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P., McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C., Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae."
RL Nature 406:477-483(2000).
RN [7]
RN SEQUENCE OF 22-124.
RX MEDLINE=78005537; PubMed=903363;
RA Kurosky A., Markel D.E., Peterson J.W.;
RT "Covalent structure of the beta chain of cholera enterotoxin."
RL J. Biol. Chem. 252:7257-7264(1977).
RN [8]
RN SEQUENCE OF 22-124.
RX MEDLINE=78005536; PubMed=903362;
RA Lai C.-Y.;
RT "Determination of the primary structure of cholera toxin B subunit."
RL J. Biol. Chem. 252:7249-7256(1977).
RN [9]
RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX MEDLINE=94272319; PubMed=8003954;
RA Merritt E.A., Sarfaty S., van den Akker F., L'Hoir C., Martial J.A., Hol W.G.J.;
RT "Crystal structure of cholera toxin B-pentamer bound to receptor GM1 pentasaccharide."
RL Protein Sci. 3:166-175(1994).
RN [10]
RN X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RX MEDLINE=95387394; PubMed=7658472;
RA Zhang R.G., Westbrook M.L., Westbrook E.M., Scott D.L., Otwinowski Z., Maulik P.R., Reed R.A., Shipley G.G.;
RT "The 2.4 A crystal structure of cholera toxin B subunit pentamer: choleragenoid."
RL J. Mol. Biol. 251:550-562(1995).
RN [11]
RN X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RC STRAIN=OGAWA 41 / CLASSICAL BIOTYPE;
RX MEDLINE=97376625; PubMed=9232653;
RA Merritt E.A., Sarfaty S., Jobling M.G., Chang T., Holmes R.K., Hirst T.R., Hol W.G.J.;
RT "Structural studies of receptor binding by cholera toxin mutants."
RL Protein Sci. 6:1516-1528(1997).
CC -!- FUNCTION: THE BETA CHAIN AGGREGATE (B SUBUNIT) IS INVOLVED IN BINDING TO CELL MEMBRANES.
CC -!- SUBUNIT: CONTAINS 3 KINDS OF CHAINS. AN ALPHA AND A GAMMA CHAIN (FROM THE SAME PRECURSOR MOLECULE), LINKED BY AN INTERCHAIN DISULFIDE BOND, ASSOCIATE NONCOVALENTLY WITH AN AGGREGATE OF 4 TO 6 BETA CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X00171; CAA24996.1; -
DR EMBL; K01170; AAA27573.1; -
DR EMBL; D30053; BAA06291.1; -
DR EMBL; X58786; CAA41593.1; -
DR EMBL; X76390; CAA53973.1; -
DR EMBL; X76391; CAA53976.1; -
DR EMBL; AE004224; AAF94613.1; -
DR PIR; A01819; XVVCB.

```

```
DR PIR: A05130; A05130.
DR PIR: S14624; S14624.
DR PDB: 2CHB; 03-DEC-97.
DR PDB: 3CHB; 12-AUG-98.
DR PDB: 1CHP; 08-MAR-96.
DR PDB: 1CHQ; 08-MAR-96.
DR PDB: 1FGB; 23-DEC-96.
DR PDB: 1XPB; 01-APR-97.
DR PDB: 1XTC; 01-AUG-96.
DR PDB: 1CT1; 15-OCT-97.
DR TIGR: VC1456; -.
DR InterPro: IPR001835; -.
DR Pfam: PF01376; Enterotoxin_B.1.
DR PRINTS: PR00772; ENTEROTOXINB.
KW Membrane; Enterotoxin; Signal; 3D-structure.
FT SIGNAL 1 21
FT CHAIN 22 124
FT DISULFID 30 107
FT CONFLICT 33 33
FT CONFLICT 39 39
FT CONFLICT 43 43
FT CONFLICT 68 68
FT CONFLICT 70 70
FT CONFLICT 75 75
FT CONFLICT 91 91
FT HELIX 26 30
FT TURN 31 32
FT TURN 34 35
FT STRAND 36 44
FT STRAND 47 51
FT TURN 54 55
FT STRAND 58 62
FT TURN 64 65
FT STRAND 68 71
FT TURN 76 77
FT HELIX 80 99
FT TURN 100 100
FT STRAND 102 109
FT STRAND 115 123
SQ SEQUENCE 124 AA; 13957 MW; 9AA393E3EA8E3EBF CRC64;

Query Match 94.3%; Score 100; DB 1; Length 124;
Best Local Similarity 95.2%; Pred. No. 1.9e-09;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GATQVEVPGSQHDSQKAI 21
Db 66 GATQVEVPGSQHDSQKAI 86

RESULT 4
IRF6_HUMAN
ID IRF6_HUMAN STANDARD; PRT; 467 AA.
AC O14896;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE INTERFERON REGULATORY FACTOR 6 (IRF-6).
GN IRF6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Grossman A., Mitrutruker H.W., Ozato K., Mak T.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Grafham D.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
```

```
CC CC -!- SIMILARITY: BELONGS TO THE IRF FAMILY.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
DR EMBL; AF027292; AAB84111.1; -.
DR EMBL; AL022398; CAA18545.1; -.
DR HSSP; P23906; 1IRG.
DR InterPro: IPR001346; -.
DR Pfam: PF00605; IRF; 1.
DR PRINTS; PR00267; INTERREGFCT.
DR PROSITE; PS00601; IRF; 1.
KW Transcription regulation; DNA-binding; Nuclear protein.
FT DNA_BIND 9 111
FT SEQUENCE 467 AA; 53129 MW; 7E28F5E0F5BA4053 CRC64;

Query Match 41.5%; Score 44; DB 1; Length 467;
Best Local Similarity 41.2%; Pred. No. 11;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Qy 5 QVEVPGSQHDSQKAI 21
Db 273 QVRFPGPEHITNEKOKL 289

RESULT 5
TGT_ZYMMO
ID TGT_ZYMMO STANDARD; PRT; 385 AA.
AC P28720; Q60247;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE QUEUINE TRNA-RIBOSYLTRANSFERASE (EC 2.4.2.29) (TRNA-GUANINE
DE TRANSGLYCOSYLASE) (GUANINE INSERTION ENZYME).
GN Tgt.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;
OC Zymomonas.
OX NCBI_TaxID=542;
RN [1]
RP SEQUENCE FROM N.A., SEQUENCE OF 1-6, AND CHARACTERIZATION.
RX MEDLINE-95394847; PubMed-7665516;
RA Reuter K.K.H., Ficner R.;
RT "Sequence analysis and overexpression of the Zymomonas mobilis tgt
RT gene encoding TRNA-guanine transglycosylase: purification and
RT biochemical characterization of the enzyme.";
RL J. Bacteriol. 177:5284-5288(1995).
RN [2]
RP PARTIAL SEQUENCE FROM N.A.
RC STRAIN-ATCC 31821 / ZM4 / Cp4;
RX MEDLINE-92406015; PubMed-15266462;
RA Shark K.B., Conway T.;
RT "Cloning and molecular characterization of the DNA ligase gene (lig)
RT from Zymomonas mobilis.";
RL FEMS Microbiol. Lett. 75:19-26(1992).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS).
RX MEDLINE-96256303; PubMed-8654383;
RA Romier C., Reuter K., Suck D., Ficner R.;
RT "Crystal structure of TRNA-guanine transglycosylase: RNA modification
RT by base exchange.";
RL EMBO J. 15:2850-2857(1996).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).
RA Graedler U., Gerber H.D., Goodenough-Lashua D.M., Garcia G.A.G.,
RA Ficner R., Reuter K., Stubbs M.T., Klebe G.;
RT "A new target for shigellosis: rational design and crystallographic
```

MEDLINE=92114793; PubMed=1837329;  
Jones D.G.L., Reusser U., Braus G.H.;  
"Molecular cloning, characterization and analysis of the regulation  
of the ARO2 gene, encoding chorismate synthase, of *Saccharomyces  
cerevisiae*.";  
Mol. Microbiol. 5:2143-2152(1991).  
[2]  
SEQUENCE FROM N.A.  
STRAIN=S288C / FY1769;  
MEDLINE=971197983; PubMed=9046099;  
Voet M., Defoor E., Verhasselt P., Riles L., Robben J., Volckaert G.;  
"The sequence of a nearly unclonable 22.8 kb segment on the left arm  
chromosome VII from *Saccharomyces cerevisiae* reveals ARO2, RPL9A,  
TIP1, MRF1 genes and six new open reading frames.";  
Yeast 13:177-182(1997).  
-1- CATALYTIC ACTIVITY: 5-O-(1-CARBOXYVINYL)-3-PHOSPHOKIMATE =

```

CC      -|  COFACTOR: REDUCED FLAVIN.
CC      -|  PATHWAY: SEVENTH STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN
CC      -|  THE BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).
CC      -|  SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC      -|  INDUCTION: BY AMINO ACID STARVATION.
CC      -|  SIMILARITY: BELONGS TO THE CHORISMATE SYNTHASE FAMILY.
CC      -----

```

between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
CC EMBL; X60190; CAA42745.1; -  
CC EMBL; X99960; CAA68214.1; -  
CC EMBL; 272670; CAA96860.1; -  
CC PIR; S17246; S17246.  
CC SGD; S0003116; ARO2.  
CC InterPro; IPR000453; -  
CC Pfam; PF01364; Chorismate\_synt; 1.  
CC PROSITE; PS00787; CHORISMATE\_SYNTHASE\_1; 1.  
CC PROSITE; PS00788; CHORISMATE\_SYNTHASE\_2; 1.  
CC PROSITE; PS00789; CHORISMATE\_SYNTHASE\_3; 1.  
KW Lyase; Aromatic amino acid biosynthesis.  
KW SEQUENCE 376 AA; .40838 MW; AF3AF65605B91E8E CRC64;  
SQ

```

Query Match      40.1%;   Score 42.5;   DB 1;   Length 376;
Best Local Similarity 62.5%;   Pred. No. 16;
Matches 10;   Conservative 2;   Mismatches 3;   Indels 1;   Gaps 1;

QY      1  GATFQ-VEVPGSQHID 15
      |:| | | | | | |
Db      258  GSGFGVSVPGSKHND 273

```

	RESULT
STI4_HUMAN	
ID	STI4_HUMAN STANDARD; PRT; 855 AA.
AC	Q9Y5Y6;
DC	01-OCT-2000 (Rel. 40, Created)
DT	01-OCT-2000 (Rel. 40, Last sequence update)
DD	01-OCT-2000 (Rel. 40, Last annotation update)
DE	SUPPRESSOR OF TUMORIGENICITY 14 (EC 3.4.21.-) (MATRIPITASE), (MEMBRANE-
DEFN	TYPE SERINE PROTEASE 1) (MT-SPI).
GN	STI4 OR PRSS14 OR SMC19.
OS	Homo sapiens (Human)
OCC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheraia; Primates; Catarrhini; Hominoidea; Homo.
NCBII_TaxID=9606;	
{1}	SEQUENCE FROM N.A. MEDLINE=99303581; PubMed=10373424; Lin C.Y., Anders J., Johnson M., Sang O.A., Dickson R.B.;



```
DR EMBL: L32961; AA74449.1; --
DR MIM: 137150; --
DR InterPro: IPR000954; --
DR Pfam: PF00202; aminotran_3; 1.
DR PROSITE: PS00600; AA_TRANSFER_CLASS_3; 1.
DR Transferrase; Amino transferase; Pyridoxal phosphate;
KW Neurotransmitter degradation; Mitochondrion; Transit peptide;
KW Disease mutation.
FT TRANSIT 1 28 MITOCHONDRION.
FT CHAIN 29 500 4-AMINO BUTYRATE AMINOTRANSFERASE.
FT BINDING 357 357 PYRIDOXAL PHOSPHATE.
FT VARIANT 220 220 R -> K (IN GABA-AT DEFICIENCY; 25%
FT REDUCTION IN ACTIVITY).
FT /FTID=VAR_008883.
FT CONFLICT 109 109 D -> H (IN REF. 2).
FT CONFLICT 113 113 V -> L (IN REF. 2).
FT CONFLICT 132 132 E -> G (IN REF. 2).
FT CONFLICT 191 191 K -> Q (IN REF. 2).
FT CONFLICT 204 204 W -> G (IN REF. 2).
FT CONFLICT 216 216 S -> A (IN REF. 2).
FT CONFLICT 268 268 G -> R (IN REF. 2).
FT CONFLICT 320 320 C -> G (IN REF. 2).
FT CONFLICT 366 366 L -> H (IN REF. 2).
FT SEQUENCE 500 AA; 56557 MW; 41199085693F80AD CRC64;

Query Match 39.2%; Score 41.5; DB 1; Length 500;
Best Local Similarity 55.6%; Pred. No. 31;
Matches 10; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 3 TFQVEVPGSQHDSQKKA 20
DB 17 TYRLVPGSRHI-SQAAA 33

RESULT 9
PULA_KLEPN STANDARD; PRT; 1090 AA.
ID PULA_KLEPN STANDARD; PRT; 1090 AA.
AC P07206;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PULLULANASE PRECURSOR (EC 3.2.1.41) (ALPHA-DEXTRIN ENDO-1,6-ALPHA-
DE GLUCOSIDASE) (PULLULAN 6-GLUCANOHYDROLASE).
GN PULA.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UNF 5023;
RX MEDLINE=90205629; PubMed=2181242;
RA Kornacker M.G., Pugsley A.P.;
RT "Molecular characterization of pula and its product, pullulanase, a
RT secreted enzyme of Klebsiella pneumoniae UNF5023."
RL Mol. Microbiol. 4:73-85(1990).
RN [2]
RP SEQUENCE OF 1-62 FROM N.A.
RX MEDLINE=86033621; PubMed=3902792;
RA Chapon C., Raibaud O.;
RT "Structure of two divergent promoters located in front of the gene
RT encoding pullulanase in Klebsiella pneumoniae and positively
RT regulated by the malt product."
RL J. Bacteriol. 164:639-645(1985).
RN [3]
RP SEQUENCE OF 944-1090 FROM N.A.
RX MEDLINE=89291709; PubMed=2661532;
RA D'Enfert C., Pugsley A.P.;
RT "Klebsiella pneumoniae pULS gene encodes an outer membrane
RT lipoprotein required for pullulanase secretion."
RL J. Bacteriol. 171:3673-3679(1989).
RL J. Bacteriol. 171:3673-3679(1989).
CC -|- CATALYTIC ACTIVITY: STARCH-DEBRANCHING ENZYME, HYDROLYSES

(1-6)-ALPHA-GLUCOSIDIC LINKAGES IN PULLULAN AND STARCH TO
FORM MALTOSE.
-|- SUBUNIT: HOMOTRIMER.
-|- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
ANCHOR (PROBABLY).
-|- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
KNOWN AS THE ALPHA-AMYLASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: X52181; CA36431.1; --
DR EMBL: M12503; AA25087.2; --
DR EMBL: M29097; AA61976.1; --
DR PIR: A25025; A25025.
DR PIR: A32880; A32880.
DR PIR: S11823; S11823.
DR InterPro: IPR000461; --
DR Pfam: PF00128; alpha-amylase; 1.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hydrolyase; Glycosidase; Membrane; Lipoprotein; Signal.
FT SIGNAL 1 19
FT CHAIN 20 1090 PULLULANASE.
FT LIPID 20 20 N-ACYL DIGLYCERIDE.
FT ACT_SITE 684 684 BY SIMILARITY.
FT ACT_SITE 713 713 BY SIMILARITY.
FT ACT_SITE 841 841 BY SIMILARITY.
FT CONFLICT 6 6 R -> C (IN REF. 2).
FT CONFLICT 10 10 V -> F (IN REF. 2).
FT CONFLICT 15 15 V -> I (IN REF. 2).
FT CONFLICT 23 23 G -> S (IN REF. 2).
FT CONFLICT 31 31 N -> S (IN REF. 2).
FT CONFLICT 34 34 T -> N (IN REF. 2).
FT CONFLICT 36 36 D -> DGNP (IN REF. 2).
FT CONFLICT 55 58 TAVE -> MATA (IN REF. 2).
FT SEQUENCE 1090 AA; 118098 MW; 240AE7DFB3FF1BD6 CRC64;

Query Match 39.2%; Score 41.5; DB 1; Length 1090;
Best Local Similarity 38.5%; Pred. No. 72;
Matches 10; Conservative 4; Mismatches 7; Indels 5; Gaps 1;

QY 1 GATFQVEVPGSQHID-----SOKKAI 21
DB 312 GVTFRVWAPTAAQOVVVYSADKKVI 337

RESULT 10
UBIQ_DICDI
ID UBIQ_DICDI STANDARD; PRT; 76 AA.
AC P08618;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-NOV-1991 (Rel. 20, Last annotation update)
DE UBIQUITIN.
OS Dictyostellium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88152253; PubMed=2831095;
RA Mueller-Taubenberger A., Westphal M., Jaeger E., Noegel A.,
RA Gerisch G.;
RT "Complete cDNA sequence of a Dictyostelium ubiquitin with a carboxy-
RT terminal tail and identification of the protein using an anti-peptide
RT antibody."
RL FEBS Lett. 229:273-278(1988).
RN [2]
```

SEQUENCE FROM N.A.  
RX MEDLINE=89352609; PubMed=2548604;  
RA Omachi T., Giorla R., Shaw D.R., Ennis H.L.;  
RT "Molecular organization of developmentally regulated Dictyostelium  
discoideum ubiquitin cDNAs.";   
RL Biochemistry 28:5226-5231(1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87257921; PubMed=3037345;  
RA Giorla R., Ennis H.L.;  
RT "Structure of two developmentally regulated Dictyostelium discoideum  
ubiquitin genes.";   
RL Mol. Cell. Biol. 7:2097-2103(1987).  
RN [4]  
RP SEQUENCE OF 13-76 FROM N.A.  
RC STRAIN=AX2;  
RA Westphal M., Mueller-Taubenberg A., Noegel A., Gerisch G.;  
RT "Transcript regulation and carboxyterminal extension of ubiquitin in  
Dictyostelium discoideum.";   
RL FEBS Lett. 209:92-96(1986).  
CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF  
CELLULAR PROTEINS. THE MAINTENANCE OF CHROMATIN STRUCTURE. THE  
REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
BIOGENESIS.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.  
CC -1- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN  
PRECURSOR WITH EXACT HEAD TO TAIL REPEATS. SOME UBIQUITIN GENES  
CONTAIN A SINGLE COPY OF UBIQUITIN FUSED TO A RIBOSOMAL PROTEIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X07210; CAA30183.1; ALT\_TERM.  
DR EMBL; M19666; AAA33261.1; ALT\_TERM.  
DR EMBL; M19491; AAA33269.1; ALT\_TERM.  
DR EMBL; M19492; AAA33270.1; ALT\_TERM.  
DR EMBL; M23748; AAA33262.1; ALT\_TERM.  
DR EMBL; M23749; AAA33263.1; ALT\_TERM.  
DR EMBL; M23750; AAA33264.1; ALT\_TERM.  
DR EMBL; M23751; AAA33265.1; ALT\_TERM.  
DR EMBL; M23752; AAA33266.1; ALT\_TERM.  
DR EMBL; M23753; AAA33267.1; ALT\_TERM.  
DR EMBL; M23754; AAA33268.1; ALT\_TERM.  
DR EMBL; X04702; CAA28408.1; ALT\_TERM.  
DR PIR; S00357; UQDOR.  
DR PIR; E34080; UQDOR7.  
DR PIR; A27806; A27806.  
DR PIR; B27806; B27806.  
DR PIR; A34080; A34080.  
DR PIR; B34080; B34080.  
DR PIR; C34080; C34080.  
DR PIR; D34080; D34080.  
DR HSP; P02248; 1AAR.  
DR DictyDb; DD05001; ubqA.  
DR InterPro; IPR000626; -.  
DR Pfam; PF00240; ubiquitin; 1.  
DR PRINTS; PR00348; UBIQUITIN.  
DR PROSITE; PS00299; UBIQUITIN\_1; 1.  
DR PROSITE; PS00553; UBIQUITIN\_2; 1.  
KW Nuclear protein; Polyprotein.  
FT SITE 48 48 NECESSARY FOR BRANCHED-CHAIN  
FT BINDING 76 76 MULTIUBIQUITIN ADDUCTS.  
FT VARIANT 11 11 K -> N (IN SOME CLONES REPEATS).  
SQ SEQUENCE 76 AA; 8538 MW; 6427383968EA8A84 CRC64;  
Query Match 38.7%; Score 41; DB 1; Length 76;

Best Local Similarity 42.9%; Pred. No. 4.9;  
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;  
Qy 1 GATFOVEVPGSQHIDSOKKAI 21  
Db 10 GKTITLEVEGSDNIENVRKI 30  
RESULT 11  
YP68\_CAEEL  
ID YP68\_CAEEL STANDARD; PRT; 313 AA.  
AC Q09217;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE HYPOTHEICAL 37.0 KDA PROTEIN B0495.8 IN CHROMOSOME II.  
GN B0495.8.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL NZ;  
RA Kirsten J.;  
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: SOME, TO YEAST YDLO87C AND S.POMBE SPCC16A11.13.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U21317; AAA62527.1; -.  
DR WormPep; B0495.8; CE01766.  
KW Hypothetical protein.  
SQ SEQUENCE 313 AA; 36977 MW; 000D2327621BFED0 CRC64;  
Query Match 38.7%; Score 41; DB 1; Length 313;  
Best Local Similarity 46.2%; Pred. No. 23;  
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;  
Qy 7 EVPQSQHIDSOKK 19  
Db 11 QLMSGSHVDNKEK 23  
RESULT 12  
YAAN\_BACSU  
ID YAAN\_BACSU STANDARD; PRT; 386 AA.  
AC P37535;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE HYPOTHEICAL 43.8 KDA PROTEIN IN XPAC-ABRB INTERGENIC REGION.  
GN YAAN.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; PubMed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
subtilis chromosome containing the replication origin.";   
RL DNA Res. 1:1-14(1994).  
RN [2]



Query Match 37.7%; Score 40; DB 1; Length 288;  
Best Local Similarity 35.3%; Pred. No. 30;  
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 3 TFQVEVPGSQHDSOKK 19  
DB 133 TFEIDVNGILHVSADK 149

RESULT 15  
ID YRVL\_CAEEL STANDARD; PRT; 355 AA.  
AC Q10005;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL 39.9 KDA PROTEIN T15H9.1 IN CHROMOSOME II PRECURSOR.  
OS T15H9.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Gardner A.;  
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: CONTAINS 1 J DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL; 247356; CAA87414.1; -  
DR HSSP; P08622; LXBL.  
DR WORMPEP; T15H9.1; CE01664.  
DR InterPro; IPR001623; -  
DR InterPro; IPR002939; -  
DR Pfam; PF00226; DnaJ; 1.  
DR Pfam; PF01556; DnaJ\_C; 1.  
DR PROSITE; PS00636; DnaJ\_1; 1.  
DR PROSITE; PS00076; DnaJ\_2; 1.  
KW Hypothetical protein; Chaperone; Signal.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 355 HYPOTHETICAL PROTEIN T15H9.1.  
FT DOMAIN 22 91 J-DOMAIN.  
FT DOMAIN 97 103 POLY-GLY.  
FT DOMAIN 116 123 POLY-GLY.  
SQ SEQUENCE 355 AA; 39861 MW; 95FA4D8E551D9CC2 CRC64;

Query Match 37.7%; Score 40; DB 1; Length 355;  
Best Local Similarity 47.4%; Pred. No. 38;  
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 3 TFQVEVPGSQHDSOKKAI 21  
DB 320 TFDVEFPKTELSDEKAQI 338

RESULT 16  
ID YDH2\_SCHPO STANDARD; PRT; 655 AA.  
AC Q92347;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHETICAL 75.4 KDA PROTEIN C6G9.02C IN CHROMOSOME I.

GN SPAC6G9.02C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Connor R.E.;  
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: TO YEAST YJL010C.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL; 281317; CAB03604.1; -  
DR InterPro; IPR001313; -  
DR Pfam; PF00806; PUF; 5.  
KW Hypothetical protein.  
SQ SEQUENCE 655 AA; 75440 MW; 45E6C4491692DA42 CRC64;

Query Match 37.7%; Score 40; DB 1; Length 655;  
Best Local Similarity 47.4%; Pred. No. 73;  
Matches 9; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 2 ATFQVEVPGSQHDSOKKA 20  
DB 535 ATLAVSAPGSHIVDKAWKA 553

RESULT 17  
RPC\_BPPHC  
ID RPC\_BPPHC STANDARD; PRT; 683 AA.  
AC P08979;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE REPRESSOR PROTEIN C.  
GN C.  
OS Bacteriophage phi-C31.  
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;  
OC Lambda phage group.  
OX NCBI\_TaxID=10719;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-NORWICH;  
RX MEDLINE=89039715; PubMed=3185504;  
RA Sinclair R.B., Bibb M.J.;  
RT "The repressor gene (c) of the Streptomyces temperate phage phi c31:  
RT nucleotide sequence, analysis and functional cloning.";  
RL Mol. Gen. Genet. 213:269-277(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94374705; PubMed=8088546;  
RA Hartley N.M., Murphy G.O., Bruton C.J., Chater K.F.;  
RT "Sequence of the essential early region of phi C31, a temperate phage  
RT of Streptomyces spp. with unusual features in its lytic  
RT development.";  
RL Gene 147:29-40(1994).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>



CC or send an email to [license@sib.ch](mailto:license@sib.ch)).

CC EMBL; X12865; CAA31345.1; -  
 CC EMBL; X76288; CAA53911.1; -  
 DR PIR; S01433; S01433.  
 KW Transcription regulation; Repressor; DNA-binding.  
 SQ SEQUENCE 683 AA; 74077 MW; B02379D204F37D1B CRC64;

Query Match 37.7%; Score 40; DB 1; Length 683;  
 Best Local Similarity 50.0%; Pred. No. 76;  
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 GATQVEVPGSQHI 14  
 | | | | | | | | | |  
 DB 86 GRTFDTLTGADHI 99

RESULT 18  
 RRPB\_CVMA5 STANDARD; PRT; 2733 AA.  
 AC P16342;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (ORF1B).  
 GN F2.  
 OS Murine coronavirus MHV (strain A59).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
 OC Coronaviridae; Coronavirus.  
 OX NCBI\_TaxID=11142;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90245573; PubMed=2159623;  
 RA Bredenbeek P.J., Pachuk C.J., Noten A.F.H., Charite J., Luytjes W.,  
 Weiss S.R., Spaan W.J.M.;  
 RT "The primary structure and expression of the second open reading  
 frame of the polymerase gene of the coronavirus MHV-A59; a highly  
 conserved polymerase is expressed by an efficient ribosomal  
 frameshifting mechanism.";  
 RT Nucleic Acids Res. 18:1825-1832(1990).  
 RL Nucleic Acids Res. 18:1825-1832(1990).  
 CC -!- FUNCTION: THE RNA DEPENDENT RNA POLYMERASE OF CORONAVIRUSES IS  
 A MULTIFUNCTIONAL PROTEIN: IT CONTAINS THE ACTIVITIES NECESSARY  
 FOR THE TRANSCRIPTION OF NEGATIVE STRANDED RNA, LEADER RNA,  
 SUBGENOMIC MRNAs AND PROGENY VIRION RNA.  
 CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -> N PYROPHOSPHATE  
 + RNA(N).  
 CC -!- MISCELLANEOUS: THIS PROTEIN MAY BE TRANSLATED AS A 1A-1B  
 POLYPEPTIDE BY A RIBOSOMAL FRAMESHIFTING MECHANISM.  
 CC -!- SIMILARITY: TO THE AVIAN CORONAVIRUS IBV RNA POLYMERASE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib.ch)).

CC EMBL; X51939; CAA36202.1; ALT\_SEQ.  
 DR PIR; S08652; S08652.  
 DR PIR; S15760; S15760.  
 KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.  
 FT DOMAIN 622 824 POLYMERASE.  
 FT DOMAIN 944 1014 CYS/HIS-RICH.  
 FT DOMAIN 1218 1506 HELICASE.  
 FT NP\_BIND 1220 1227 ATP (BY SIMILARITY).  
 SQ SEQUENCE 2733 AA; 309218 MW; F3ACC8EF20D20C41 CRC64;

Query Match 37.7%; Score 40; DB 1; Length 2733;  
 Best Local Similarity 50.0%; Pred. No. 3.4e+02;  
 Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 TQQVEVPGSQHDSQX 18  
 | | | | | | | | | |  
 DB 1199 TFGNNVPNYQHGMKR 1214

RESULT 19  
 INRL\_SHEEP STANDARD; PRT; 560 AA.  
 ID INRL\_SHEEP Q95206;  
 AC Q28589; Q95206;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE INTERFERON-ALPHA/BETA RECEPTOR ALPHA CHAIN PRECURSOR (IFN-ALPHA-REC)  
 DE (INTERFERON ALPHA/BETA RECEPTOR-1).  
 OS IFNARI OR IFNAR.  
 GN Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Caprinae; Ovis.  
 OX NCBI\_TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Endometrium;  
 RX MEDLINE=97135690; PubMed=8981227;  
 RA Kaluz S., Fisher P.A., Kaluzova M., Sheldrick E.L., Flint A.P.F.;  
 RT "Structure of an ovine interferon receptor and its expression in  
 endometrium.";  
 RT J. Mol. Endocrinol. 17:207-215(1996).  
 RL J. Mol. Endocrinol. 17:207-215(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Endometrium;  
 RX MEDLINE=98006426; PubMed=9348203;  
 RA Han C.-S., Mathialagan N., Klemann S.W., Roberts R.M.;  
 RT "Molecular cloning of ovine and bovine type I interferon receptor  
 subunits from uteri, and endometrial expression of messenger  
 ribonucleic acid for ovine receptors during the estrous cycle and  
 pregnancy.";  
 RT Endocrinology 138:4757-4767(1997).

CC -!- FUNCTION: RECEPTOR FOR INTERFERONS ALPHA AND BETA. BINDING TO TYPE  
 I IFNS TRIGGERS TYROSINE PHOSPHORYLATION OF A NUMBER OF PROTEINS  
 INCLUDING JAKS, TYK2, STAT PROTEINS AND IFN-R ALPHA-AND BETA-  
 SUBUNIT'S THEMSELVES.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXAMINED EXCEPT  
 CONCEPTUS AT DAY 15 OF PREGNANCY.  
 CC -!- SIMILARITY: BELONGS TO THE CLASS II CYTOKINE FAMILY OF RECEPTORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL; X95939; CAA65183.1; -  
 DR EMBL; U65978; AAB84231.1; -  
 DR InterPro; IPR001777; -  
 DR Pfam; PF00041; fn3; 1.  
 KW Receptor; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 24 BY SIMILARITY.  
 FT CHAIN 25 560 INTERFERON-ALPHA/BETA RECEPTOR ALPHA  
 CHAIN.  
 FT DOMAIN 25 437 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 438 458 POTENTIAL.  
 FT DOMAIN 459 560 CYTOPLASMIC (POTENTIAL).  
 FT DISULFID 76 84 BY SIMILARITY.  
 FT DISULFID 199 220 BY SIMILARITY.  
 FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 55 55 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).

```
FT CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 377 377 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 S -> G (IN REF. 2).
FT CONFLICT 522 522 A -> D (IN REF. 2).
SQ SEQUENCE 560 AA; 63918 MW; E7198A1905D4805C CRC64;

Query Match 37.3%; Score 39.5; DB 1; Length 560;
Best Local Similarity 37.5%; Pred. No. 74;
Matches 9; Conservative 4; Mismatches 4; Indels 7; Gaps 1;

Oy 2 ATFOV-----EVPQSQHSQKAI 18
| :| :||| | | | | |
Db 60 ADYQILGTDNWKLPQCQHITSSK 83
| :| :||| | | | | |

RESULT 20
Y13K_SSV1
ID Y13K_SSV1 STANDARD; PRT; 112 AA.
AC P20220;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE HYPOTHEICAL 13.2 KDA PROTEIN (ORF F-112).
OS Sulfolobus virus-like particle SSV1.
OC Viruses; dsDNA viruses, no RNA stage; Fuselloviridae; Fusellovirus.
OX NCBI_TaxID=10476;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92024080; PubMed=1926776;
RA Palm P., Schleper C., Gramp B., Yeats S., McWilliam P., Reiter W.-D.,
RA Zillig W.;
RT "Complete nucleotide sequence of the virus SSV1 of the
RT Archaeobacterium Sulfolobus shibatae.";
RL Virology 185:242-250(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X07234; CAA30220.1; -.
DR PIR; S03221; S03221.
KW Hypothetical protein.
SQ SEQUENCE 112 AA; 13204 MW; 23F9C02D979AF383 CRC64;

Query Match 36.8%; Score 39; DB 1; Length 112;
Best Local Similarity 40.0%; Pred. No. 16;
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Oy 2 ATFOVEVPGSQHSQKAI 21
| :| :| :| | | | |
Db 31 AQFELVPSAVNIQALKAI 50
| :| :| :| | | | |

RESULT 21
YC43_ODOSI
ID YC43_ODOSI STANDARD; PRT; 263 AA.
AC P49538;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
```

```
DE HYPOTHEICAL 30.1 KDA PROTEIN YCF43 (ORF263).
GN YCF43.
OS Odontella sinensis.
OC Chloroplast.
OC Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
OC Biddulphiophycidae; Eupodiscales; Eupodisaceae; Odontella.
OX NCBI_TaxID=2839;
RN [1]
RP SEQUENCE FROM N.A.
RA Kowalik K.V., Stoebe B., Schaffran I., Kroth-Pancic P., Frieler U.;
RT "The chloroplast genome of a chlorophyll a+c-containing alga,
RT Odontella sinensis.";
RL Plant Mol. Biol. Rep. 13:336-342(1995).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE TATC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z67753; CAA91737.1; -.
DR Mendel; 5328; ODOI:ycf43.1.
DR InterPro; IPR002033; -.
DR Pfam; PF00902; UPF0032; 1.
DR PROSITE; PS01218; TATC; 1.
KW Chloroplast; Hypothetical protein; Transmembrane.
FT TRANSMEM 53 73 POTENTIAL.
FT TRANSMEM 103 123 POTENTIAL.
FT TRANSMEM 130 150 POTENTIAL.
FT TRANSMEM 153 173 POTENTIAL.
FT TRANSMEM 181 201 POTENTIAL.
FT TRANSMEM 213 233 POTENTIAL.
FT TRANSMEM 241 261 POTENTIAL.
SQ SEQUENCE 263 AA; 30080 MW; 25083364F2A89E88 CRC64;

Query Match 36.8%; Score 39; DB 1; Length 263;
Best Local Similarity 36.8%; Pred. No. 40;
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Oy 3 TFQVEVPGSQHSQKAI 21
| :| :| :| | | | |
Db 31 TVTLELPFSEHIELKQRL 49
| :| :| :| | | | |

RESULT 22
SR54_THEAC
ID SR54_THEAC STANDARD; PRT; 456 AA.
AC Q9HKT0;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE SIGNAL RECOGNITION 54 KDA PROTEIN (SRP54).
GN SRP54 OR TA0515.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;
OC Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum";
RL Nature 407:508-513(2000).
CC -!- FUNCTION: BINDS TO THE SIGNAL SEQUENCE OF PRESECRETORY PROTEIN
CC WHEN THEY EMERGE FROM THE RIBOSOMES (BY SIMILARITY).
```

CC -!- SUBUNIT: ARCHAEAL SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA  
CC MOLECULE OF 300 NUCLEOTIDES AND TWO PROTEIN SUBUNITS: SRP54 AND  
CC SRP19 (BY SIMILARITY).  
CC  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
CC  
CC -!- DOMAIN: HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN BINDS GTP; THE  
CC M-DOMAIN BINDS THE 7S RNA IN PRESENCE OF SRP19 AND ALSO BINDS THE  
CC SIGNAL SEQUENCE (BY SIMILARITY).  
CC  
CC -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AL445064; CAC11655.1; .  
CC DR PROSITE; P500300; SRP54; 1.  
CC KW Signal recognition particle; GTP-binding; RNA-binding.  
CC FT DOMAIN 1 289 G-DOMAIN (BY SIMILARITY).  
CC FT NP\_BIND 104 111 M-DOMAIN (BY SIMILARITY).  
CC FT NP\_BIND 184 188 GTP (BY SIMILARITY).  
CC FT NP\_BIND 242 245 GTP (BY SIMILARITY).  
CC SQ SEQUENCE 456 AA; 51080 MW; E6560FA972DA3EA7 CRC64;  
CC  
CC Query Match 36.8%; Score 39; DB 1; Length 456;  
CC Best Local Similarity 63.6%; Pred. No. 72;  
CC Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
CC  
CC QY 8 VPGSQHDSQK 18  
CC :|||||  
CC DB 353 IPGSQKIDDSK 363  
CC  
CC RESULT 23  
CC IL7R\_MOUSE  
CC ID IL7R\_MOUSE STANDARD; PRT; 459 AA.  
CC AC P16872;  
CC DT 01-AUG-1990 (Rel. 15, Created)  
CC DT 01-AUG-1990 (Rel. 15, Last sequence update)  
CC DT 15-DEC-1998 (Rel. 37, Last annotation update)  
CC DE INTERLEUKIN-7 RECEPTOR ALPHA CHAIN PRECURSOR (IL-7R-ALPHA).  
CC GN IL7R.  
CC OS Mus musculus (Mouse).  
CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
CC OX NCBI\_TaxID=10090;  
CC [1]  
CC SEQUENCE FROM N.A.  
CC RP MEDLINE-90199875; PubMed-2317865;  
CC RA Goodwin R.G., Friend D., Ziegler S.F., Jerzy R., Falk B.A.,  
CC RA Gimpel S., Cosman D., Dower S.K., March C.J., Namen A.E., Park L.S.;  
CC RT "Cloning of the human and murine interleukin-7 receptors:  
CC RT demonstration of a soluble form and homology to a new receptor  
CC RT superfamily.";  
CC RL Cell 60:941-951(1990).  
CC -!- FUNCTION: RECEPTOR FOR INTERLEUKIN-7.  
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA CHAIN AND A COMMON GAMMA CHAIN.  
CC -!- TISSUE SPECIFICITY: SPLEEN; THYMUS; AND FETAL LIVER.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -!- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.  
CC -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

CC EMBL: M29697; AAA39304.1; .  
CC DR FIR; D34791; D34791.  
CC DR MGI: 96562; IL7r.  
CC DR InterPro; IPR000950; .  
CC DR InterPro; IPR001777; .  
CC DR InterPro; IPR002465; .  
CC DR Pfam; PF00041; f03; 1.  
CC DR PROSITE; P501355; HEMATOPO\_REC\_S\_F1; 1.  
CC KW Receptor; Transmembrane; Glycoprotein; Phosphorylation; Signal.  
CC FT SIGNAL 1 20  
CC FT CHAIN 21 459 INTERLEUKIN-7 RECEPTOR ALPHA CHAIN.  
CC FT DOMAIN 21 239 EXTRACELLULAR (POTENTIAL).  
CC FT TRANSMEM 240 264 POTENTIAL.  
CC FT DOMAIN 265 459 CYTOPLASMIC (POTENTIAL).  
CC FT DOMAIN 128 225 FIBRONECTIN TYPE-III.  
CC FT DOMAIN 184 189 SER/THR-RICH.  
CC FT CARBOHYD 60 60 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 115 115 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 177 177 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT MOD\_RES 282 282 PHOSPHORYLATION (BY PKC) (POTENTIAL).  
CC SQ SEQUENCE 459 AA; 51704 MW; CC06A5CE95543849 CRC64;  
CC  
CC Query Match 36.8%; Score 39; DB 1; Length 459;  
CC Best Local Similarity 70.0%; Pred. No. 72;  
CC Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
CC  
CC QY 5 QVEVPGSQHI 14  
CC :|||||  
CC DB 45 QLEVDGSQL 54  
CC  
CC RESULT 24  
CC CD5\_RAT  
CC ID CD5\_RAT STANDARD; PRT; 491 AA.  
CC AC P5182;  
CC DT 01-OCT-1996 (Rel. 34, Created)  
CC DT 01-OCT-1996 (Rel. 34, Last sequence update)  
CC DT 30-MAY-2000 (Rel. 39, Last annotation update)  
CC DE T-CELL SURFACE GLYCOPROTEIN CD5 PRECURSOR (LYMPHOCYTE GLYCOPROTEIN LY-  
CC DE 1) (LYT-1) (LYMPHOCYTE ANTIGEN CD5).  
CC GN CD5  
CC OS Rattus norvegicus (Rat).  
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
CC OX NCBI\_TaxID=10116;  
CC [1]  
CC SEQUENCE FROM N.A.  
CC RA Murakami T.;  
CC RL Submitted (JUL-1992) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: MAY ACT AS A RECEPTOR IN REGULATING T-CELL  
CC PROLIFERATION. CD5 INTERACTS WITH CD72/LYB-2.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -!- SIMILARITY: CONTAINS 2 SRCR DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: D10728; BAA01571.1; .  
CC DR InterPro; IPR001190; .  
CC DR Pfam; PF00530; SRCR; 3.  
CC DR PROSITE; P500420; SRCR\_1; FALSE\_NEG.  
CC KW Signal; Transmembrane; Glycoprotein; T-cell.  
CC FT SIGNAL 1 23 BY SIMILARITY.  
CC FT CHAIN 24 491 T-CELL SURFACE GLYCOPROTEIN CD5.  
CC FT DOMAIN 25 368 EXTRACELLULAR (POTENTIAL).  
CC FT TRANSMEM 369

FT DOMAIN 399 491 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 33 131 SRCR 1.  
FT DOMAIN 273 378 SRCR 2.  
FT DISULFID 43 105 POTENTIAL.  
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 176 176 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 491 AA; 53440 MW; 88DD76F856891CDE CRC64;

Query Match 36.8%; Score 39; DB 1; Length 491;  
Best Local Similarity 46.2%; Pred. No. 78;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 5 QVEVPGSOHDSQ 17  
||| | : | : | :  
DB 436 QVENPAASHVDNE 448

## RESULT 25

ID HS7C\_CAEEL STANDARD; PRT; 661 AA.  
AC P27420;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE HEAT SHOCK 70 KDA PROTEIN C PRECURSOR.  
GN HSP-3 OR HSP70C.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89356251; PubMed=2766926;  
RA Heschl M.F.P.; Baillie D.L.;  
RT "Characterization of the hsp70 multigene family of Caenorhabditis elegans."  
RL DNA 8:233-243(1989).  
CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.  
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY. STRONG,  
CC TO MAMMALIAN GRP78.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; M26604; AAA28074.1; -  
PIR; A32475; A32475.  
DR HSSP; P19120; INGJ.  
DR InterPro; IPR00086; -  
DR InterPro; IPR001023; -  
DR Pfam; PF00012; HSP70; 1.  
DR PRINTS; PR00301; HEATSHOCK70.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
DR PROSITE; PS00297; HSP70\_1; 1.  
DR PROSITE; PS00329; HSP70\_2; 1.  
DR PROSITE; PS01036; HSP70\_3; 1.  
KW ATP-binding; Heat shock; Endoplasmic reticulum; Multigene family;  
KW Signal.  
FT SIGNAL 1 17 POTENTIAL.  
FT CHAIN 18 661 HEAT SHOCK 70 KDA PROTEIN C.  
FT SITE 658 661 PREVENT SECRETION FROM ER.  
SQ SEQUENCE 661 AA; 73349 MW; 7F5CD624F48293BA CRC64;

Query Match 36.8%; Score 39; DB 1; Length 661;  
Best Local Similarity 35.3%; Pred. No. 1.le+02;  
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 3 TFOVEVPGSOHDSOKK 19  
||| : | : | : | :  
DB 505 TFEIDVNGILHVTAEK 521

Search completed: July 16, 2001, 16:44:04  
Job time: 463 sec



***This Page Blank (uspto)***

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 16:43:37 ; Search time 57.41 Seconds  
(without alignments)  
48.396 Million cell updates/sec

Title: US-09-786-648-4

Perfect score: 106

Sequence: 1 GATQVEVPGSQHDSQKAI 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPREMBL\_16:\*
- 1: sp.archaea:\*
  - 2: sp.bacteria:\*
  - 3: sp.fungi:\*
  - 4: sp.human:\*
  - 5: sp.invertebrate:\*
  - 6: sp.mammal:\*
  - 7: sp.mhc:\*
  - 8: sp.organelle:\*
  - 9: sp.phage:\*
  - 10: sp.plant:\*
  - 11: sp.rodent:\*
  - 12: sp.unclassified:\*
  - 13: sp.vertebrate:\*
  - 14: sp.virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	100.0	124	2 Q57193	Q57193 vibrio chol
2	106	100.0	124	2 Q56635	Q56635 vibrio chol
3	106	100.0	124	2 Q9RP15	Q9RP15 vibrio chol
4	100	94.3	103	2 Q9R646	Q9R646 vibrio chol
5	52	49.1	552	10 Q9LJB0	Q9LJB0 arabidopsis
6	51	48.1	508	4 Q99987	Q99987 homo sapien
7	49	46.2	428	10 Q9LGM2	Q9LGM2 oryza sativ
8	48	45.3	392	10 Q40742	Q40742 oryza sativ
9	47	44.3	395	5 Q9NND5	Q9NND5 drosophila
10	46.5	43.9	1166	3 Q9PAE0	Q9PAE0 ustilago ma
11	46	43.4	293	10 Q49876	Q49876 lupinus alb
12	46	43.4	374	5 Q9U3M9	Q9U3M9 caenorhabdi
13	46	43.4	574	10 Q9SXC9	Q9SXC9 arabidopsis
14	44.5	42.0	374	2 Q9PKY2	Q9PKY2 chlamydia m
15	44	41.5	91	10 Q9SYF4	Q9SYF4 arabidopsis
16	44	41.5	330	11 Q9QZL7	Q9QZL7 mus musculus
17	44	41.5	467	6 Q9N136	Q9N136 ovis aries
18	44	41.5	467	11 P97431	P97431 mus musculus
19	44	41.5	1742	2 Q55583	Q55583 synechocyst

20	43.5	41.0	565	10 Q22511	Q22511 vitis vinif
21	43	40.6	298	10 Q43098	Q43098 psophocarpu
22	43	40.6	356	5 Q9NF90	Q9NF90 leishmania
23	43	40.6	365	1 Q27066	Q27066 methanobact
24	43	40.6	399	2 Q9F5L7	Q9F5L7 zymomonas m
25	43	40.6	439	10 Q9LY67	Q9LY67 arabidopsis
26	42	39.6	216	2 Q9K922	Q9K922 bacillus ba
27	42	39.6	298	10 Q9SXM5	Q9SXM5 glycine max
28	42	39.6	299	10 Q9S7G9	Q9S7G9 glycine max
29	42	39.6	325	4 Q9HB36	Q9HB36 homo sapien
30	42	39.6	427	13 Q42099	Q42099 cyprinus ca
31	42	39.6	787	2 Q9PKE3	Q9PKE3 chlamydia m
32	42	39.6	855	4 Q9HCA3	Q9HCA3 homo sapien
33	42	39.6	855	4 Q9H3S0	Q9H3S0 homo sapien
34	42	39.6	864	3 Q9F5Z5	Q9F5Z5 neurospora
35	41	38.7	91	2 Q9JRI4	Q9JRI4 neisseria m
36	41	38.7	246	5 Q9V9U6	Q9V9U6 drosophila
37	41	38.7	413	2 Q46961	Q46961 erwinia chr
38	41	38.7	550	10 Q04636	Q04636 arabidopsis
39	41	38.7	551	10 Q9S079	Q9S079 pinus taeda
40	41	38.7	843	14 Q9I529	Q9I529 hepatitis b
41	41	38.7	843	14 Q9QMN7	Q9QMN7 hepatitis b
42	41	38.7	2911	5 Q93442	Q93442 caenorhabdi
43	40.5	38.2	361	5 Q9VGT2	Q9VGT2 drosophila
44	40.5	38.2	366	5 Q9VLB9	Q9VLB9 drosophila
45	40.5	38.2	392	5 Q9I7M2	Q9I7M2 drosophila

ALIGNMENTS

RESULT 1

Q57193 ID Q57193 PRELIMINARY: PRT; 124 AA.

AC Q57193; DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE CHOLERA TOXIN B PROTEIN (CTB) PRECURSOR (CTB).

GN CTXB.

OS Vibrio cholerae.

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

OX NCBI\_TaxID=666;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CLASSICAL STRAIN 569B;

RX MEDLINE=9135224; PubMed=1883840;

RA Dams E., De Wolf M., Dierick W.;

RT "Nucleotide sequence analysis of the CT operon of the Vibrio cholerae

RT classical strain 569B.";

RL Biochim. Biophys. Acta 1090:139-141(1991).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CLASSICAL BIOTYPE 569B;

RA Shi C., Cao C., Zhang J., Ma Q.;

RL Chin. Biochem. J. 9:395-399(1993).

[3]

RP SEQUENCE FROM N.A.

RC STRAIN=CLASSICAL BIOTYPE 569B;

RA Xu L.;

RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.

DR EMBL; X58785; CAA41591.1; -;

DR EMBL; U25679; AAC34728.1; -;

DR EMBL; A00931; CAA00098.1; -;

DR HSSP; P01556; 2CHB.

DR InterPro; IPR001835; -;

DR Pfam; PF01376; Enterotoxin\_B.1.

DR PRINTS; PR00772; ENTEROTOXINB.

DR ProDom; PD012805; -; 1.

DR Signal.

FT SIGNAL

FT CHAIN 1 21 POTENTIAL.

FT CHAIN 22 124 CHOLERA TOXIN B PROTEIN (CTB).

SEQUENCE 124 AA; 13919 MW; D6BF83FFF7924EA3 CRC64;

Query Match 100.0%; Score 106; DB 2; Length 124;  
Best Local Similarity 100.0%; Pred. No. 2.9e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSQKAI 21  
|||||  
Db 66 GATFOVEVPGSQHIDSQKAI 86

RESULT 2  
Q56635 PRELIMINARY; PRT; 124 AA.  
AC Q56635;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
DE CHOLERA TOXIN PRECURSOR.  
GN CTXB.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S7;  
RA Yamamoto K., Do V.G., Xu M., Iida T., Miwatani T., Albert M.J.,  
Honda T.;  
RT "Comparison of cholera toxin genes (ctxAB) of non-O1 vibrio cholerae  
strains 854 (O139-bengal) and S7 (O37) from two outbreaks.";  
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.  
DR EMBL: D30052; BAA06289.1; .  
DR HSP: P01556; 2CHB.  
DR InterPro: IPR001835; .  
DR Pfam: PF01376; Enterotoxin\_B; 1.  
DR PRINTS: PR00772; ENTEROTOXINB.  
DR ProDom: PD012805; -. 1.  
KW Signal.  
FT SIGNAL.  
SQ SEQUENCE 124 AA; 13871 MW; 3F87B2F297953179 CRC64; .

Query Match 100.0%; Score 106; DB 2; Length 124;  
Best Local Similarity 100.0%; Pred. No. 2.9e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSQKAI 21  
|||||  
Db 66 GATFOVEVPGSQHIDSQKAI 86

RESULT 3  
Q56635 PRELIMINARY; PRT; 124 AA.  
AC Q56635;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
DE CHOLERA ENTEROTOXIN B-SUBUNIT.  
GN CTXB.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KNIH002;  
RA Shin H.J., Park Y.C., Kim Y.C.;  
RT "Cloning and nucleotide sequence analysis of the virulence gene  
cassette from Vibrio cholerae KNIH002 isolated in Korea.";  
RL Misaimurhag Hoiji 35:205-210(1999).  
DR EMBL: AF175708; AAD51360.1; .  
DR HSP: P01556; 2CHB.  
DR InterPro: IPR001835; .

DR Pfam: PF01376; Enterotoxin\_B; 1.  
DR PRINTS: PR00772; ENTEROTOXINB.  
SQ SEQUENCE 124 AA; 13905 MW; 23BF83FF793E5B9 CRC64;

Query Match 100.0%; Score 106; DB 2; Length 124;  
Best Local Similarity 100.0%; Pred. No. 2.9e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSQKAI 21  
|||||  
Db 66 GATFOVEVPGSQHIDSQKAI 86

RESULT 4  
Q56635 PRELIMINARY; PRT; 103 AA.  
AC Q56635;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
DE CHOLERA-LIKE ENTEROTOXIN B SUBUNIT.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=95303036; PubMed=7783690;  
RA Nakashima K., Eguchi Y., Nakasone N.;  
RT "Characterization of an enterotoxin produced by Vibrio cholerae  
O139.";  
RL Microbiol. Immunol. 39:87-94(1995).  
DR HSP: P01556; 1XTC.  
DR InterPro: IPR001835; .  
DR Pfam: PF01376; Enterotoxin\_B; 1.  
DR PRINTS: PR00772; ENTEROTOXINB.  
SQ SEQUENCE 103 AA; 11645 MW; 992A05C077F2F70E CRC64;

Query Match 94.3%; Score 100; DB 2; Length 103;  
Best Local Similarity 95.2%; Pred. No. 2.4e-09;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSQKAI 21  
|||||  
Db 45 GATFOVEVPGSQHIDSQKAI 65

RESULT 5  
Q56635 PRELIMINARY; PRT; 552 AA.  
AC Q56635;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
DE EMB|CAB77996.1.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX PubMed=10907853;  
RA Nakamura Y.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
Sequence features of the regions of 4,251,695 bp covered by ninety pl,



```

RT TAC and BAC clones.
RL DNA Res. 7:217-221(2000).
DR EMBL; AF000606; BAB01195.1; -. 740EAL6CDEBB2447 CRC64;
SQ SEQUENCE 552 AA; 63036 MW; 740EAL6CDEBB2447 CRC64;

Query Match 49.1%; Score 52; DB 10; Length 552;
Best Local Similarity 58.8%; Pred. No. 1.9;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 QVEVPGSQHIDSOKKAI 21
   |||:| | | | |
Db 147 QVEIPASMEIDEETKAI 163

RESULT 6
ID Q99987 PRELIMINARY; PRT; 508 AA.
AC Q99987;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE VRK2, COMPLETE CDS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=98008921; PubMed=9344656;
RA Nezu J., Oku A., Jones M.H., Shimane M.;
RT "Identification of two novel human putative serine/threonine kinases,
RT VRK1 and VRK2, with structural similarity to vaccinia virus B1R
RT kinase.";
RL Genomics 45:327-331(1997).
CC -!- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AB000450; BAA19109.1; -.
DR HSSP; Q06486; ICKI.
DR InterPro; IPR002719; -.
DR InterPro; IPR002290; -.
DR Pfam; PF00069; pkinase; 2.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR SMART; SM00220; S_TKC; 1.
KW ATP-binding; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 508 AA; 58126 MW; 157FBF8F48511AF4 CRC64;

Query Match 48.1%; Score 51; DB 4; Length 508;
Best Local Similarity 50.0%; Pred. No. 2.5;
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKA 20
   | : | | | | | | |
Db 329 GQSIHVHTPNSQKVDQKAA 348

RESULT 7
Q9LGM2 PRELIMINARY; PRT; 428 AA.
ID Q9LGM2;
AC Q9LGM2;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE ESTS AU056822(S20908).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
OC Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponebare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0041E11.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponebare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0433F09.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF002521; BAA96762.1; -.
DR EMBL; AF002539; BAB08201.1; -.
DR InterPro; IPR001552; -.
DR Pfam; PF00441; Acyl-CoA_dh; 1.
DR PROSITE; PS00073; ACYL_COA_DH_2; UNKNOWN_1.
SQ SEQUENCE 428 AA; 46132 MW; 8D34E3698A8E6367 CRC64;

Query Match 46.2%; Score 49; DB 10; Length 428;
Best Local Similarity 52.6%; Pred. No. 4.5;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 TFQVEVPGSQHIDSOKKAI 21
   | : | | | | | | |
Db 183 TTATKVPGGWHIDGQRWI 201

RESULT 8
Q40742 PRELIMINARY; PRT; 392 AA.
ID Q40742;
AC Q40742;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE OSRAD23
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
OC Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NIPPONBARE;
RX MEDLINE=97369378; PubMed=9225866;
RA Schultz T.F., Quatrano R.S.;
RT "Characterization and expression of a rice RAD23 gene.";
RL Plant Mol. Biol. 34:557-562(1997);
DR EMBL; U63530; AAB65841.1; -.
DR HSSP; P02248; ITBE.
DR InterPro; IPR000449; -.
DR InterPro; IPR000626; -.
DR Pfam; PF00240; ubiquitin; 1.
DR Pfam; PF00627; UBA; 2.
DR PROSITE; PS50053; UBIQUITIN_2; 1.
DR SMART; SM00165; UBA; 1.
SQ SEQUENCE 392 AA; 41753 MW; BD6E08574CC7CACB CRC64;

Query Match 45.3%; Score 48; DB 10; Length 392;
Best Local Similarity 42.9%; Pred. No. 6;
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21
   | : | | | | | | |
Db 10 GSTFQIEVDSAQKVDVKRII 30

RESULT 9
Q9NKK5 PRELIMINARY; PRT; 395 AA.
ID Q9NKK5

```

AC Q9NKG5;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE HYPOTHETICAL 43.6 KDA PROTEIN.  
GN BG:DS01514.3.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Y, CN BW SP;  
RX MEDLINE=99403001; PubMed=10471707;  
RA Ashburner M., Misra S., Rote J., Lewis S.E., Blazej R., Davis T.,  
RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,  
RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,  
RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,  
RA Celniker S., Rubin G.M.;  
RT "An exploration of the sequence of a 2.9-Mb region of the genome of  
RT Drosophila melanogaster: the Adh region.";  
RL Genetics 153:179-219(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Y, CN BW SP;  
RA Celniker S.E., Aghayani A., Arcalini T.T., Baxter E., Blazej R.G.,  
RA Butenoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,  
RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,  
RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,  
RA Lewis S., Li P., Lomoton M.A., Mazda P., Moshrefi A.R., Moshrefi M.,  
RA Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,  
RA Sethi H., Snir E., Svirska R.R., Wan K.H., Weinburg T., Zhang R.,  
RA Zierahn L.L., Rubin G.M.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DE EMBL; AE003408; AAF4847.1; -;  
DR FlyBase; FBgn0028907; BG:DS01514.3.  
KW Hypothetical protein.  
SQ SEQUENCE 395 AA; 43561 MW; AE4F1CC4ADD3DA73 CRC64;

Query Match 44.3%; Score 47; DB 5; Length 395;  
Best Local Similarity 56.2%; Pred. No. 9;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 QVEVPGSQHDSQKKA 20  
DB 329 RVSVPGSTHIDADANA 344  
| | | | | | | | | |

RESULT 10  
Q9P4E0 PRELIMINARY; PRT; 1166 AA.  
AC Q9P4E0;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE PHEROMONE-RESPONSIVE MAPKK KINASE UBC4.  
GN UBC4.  
OS Ustilago maydis (Smut fungus).  
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;  
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.  
OX NCBI\_TaxID=5270;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1/2;  
RX MEDLINE=20331594; PubMed=10875339;  
RA Andrews D.L., Egan J.D., Mayorga M.E., Gold S.E.;  
RT "The Ustilago maydis ubc4 and ubc5 genes encode members of a MAP  
RT kinase cascade required for filamentous growth.";  
RL Mol. Plant Microbe Interact. 13:781-786(2000).  
CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.  
DR EMBL; AF197562; AAF66841.1; -;

DR InterPro; IPR000104; -;  
DR InterPro; IPR000159; -;  
DR InterPro; IPR000719; -;  
DR InterPro; IPR002290; -;  
DR Pfam; PF00069; pkinase; 1.  
DR PRINTS; PRO0308; ANTI-FREEZE1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR SMART; SMO0220; S\_TKc; 1.  
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
SQ SEQUENCE 1166 AA; 123298 MW; A89E0019AABD4172 CRC64;

Query Match 43.9%; Score 46.5; DB 3; Length 1166;  
Best Local Similarity 47.6%; Pred. No. 35;  
Matches 10; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

QY 2 ATTFQVEVP-GSQHDSQKKA 21  
DB 729 AVKQVELPGSDSHLDQRKKG 749  
| | | | | | | | | |

RESULT 11  
O49876 PRELIMINARY; PRT; 293 AA.  
AC O49876;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE CLASS III CHITINASE PRECURSOR (EC 3.2.1.14) (CHITODEXTRINASE) (1,4-  
DE BETA-POLY-N-ACETYLGLUCOSAMINIDASE) (POLY-BETA-GLUCOSAMINIDASE).  
GN CHIB1 OR LUPAL.  
OS Lupinus albus (white lupine).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
OC Fabales; Fabaceae; Papilionoideae; Lupinus.  
OX NCBI\_TaxID=3870;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. ULTRA;  
RA Regalado A.P., Vidal S., Neves A., Ricardo C.P.P.,  
RA Rodrigues-Pousada C.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF N-  
CC ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.  
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL  
CC HYDROLASES).  
DR EMBL; Y16415; CAA76203.1; -;  
DR HSSP; P23472; 2HVM.  
DR Mendel; 28297; Lupal; Chib1; 28297.  
DR InterPro; IPR001579; -;  
DR Pfam; PF00192; chitinase\_2; 1.  
DR PROSITE; PS01095; CHITINASE\_18; 1.  
KW Glycosidase; Hydrolase; Signal.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 293 POTENTIAL.  
SQ SEQUENCE 293 AA; 31128 MW; DB1B21728F657F2F CRC64;

Query Match 43.4%; Score 46; DB 10; Length 293;  
Best Local Similarity 38.1%; Pred. No. 9.6;  
Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 GATFQVEVPGSQHDSQKKA 21  
DB 142 GIDFDEIAGGAQHYDELARAL 162  
| | | | | | | | | |

RESULT 12  
Q9U3M9 PRELIMINARY; PRT; 374 AA.  
ID Q9U3M9  
AC Q9U3M9;

```

DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE C40H5.3 PROTEIN.
GN C40H5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RN SEQUENCE FROM N.A.
RA White S.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: 281482; CAB03934.2; -.
SQ SEQUENCE 374 AA; 42198 MW; 67D202886D6A7824 CRC64;

Query Match 43.4%; Score 46; DB 5; Length 374;
Best Local Similarity 42.1%; Pred. No. 12;
Matches 8; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSQKK 19
| | | | | | | |
Db 230 GVEFHWFPGPHADQKE 248

RESULT 13
Q9SXC9 PRELIMINARY; PRT; 574 AA.
AC Q9SXC9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE T17H3.2 PROTEIN.
GN T17H3.2.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Lee J.M., Liu A., Li J., Gonzalez A., Liu K., Vaysberg M., Sakano H.,
RA Chin C., Hwang B., Choi E., Chlou J., Altafi H., Araujo R., Brooks S.,
RA Buehler E., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,
RA Huizar L., Khan S., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC T17H3 sequence.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC005916; AAD45990.1; -.
DR InterPro: IPR001382; -.
DR PRINTS: PR00747; GLYDRLASE47.
DR SEQUENCE 574 AA; 65707 MW; 4ACC456DE487EA93 CRC64;

Query Match 43.4%; Score 46; DB 10; Length 574;
Best Local Similarity 58.8%; Pred. No. 20;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 TFOVEVPGSQHIDSQKK 19
| | | | | | | |
Db 19 TFEVDPSSQHIETKKK 35

```

```

RESULT 14
Q9PKY2 PRELIMINARY; PRT; 374 AA.
AC Q9PKY2;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHETICAL PROTEIN TC0328.
GN TC0328.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=MOPN / NIGG;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brumham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL: AE002300; AAF39192.1; -.
DR TIGR: TC0328; -.
KW Hypothetical protein.
SQ SEQUENCE 374 AA; 41218 MW; 656B3885D9622D7F CRC64;

Query Match 42.0%; Score 44.5; DB 2; Length 374;
Best Local Similarity 48.0%; Pred. No. 22;
Matches 12; Conservative 3; Mismatches 5; Indels 5; Gaps 1;

QY 2 ATFOVEVPGSOH-----IDSQKKAI 21
| | | | | | | |
Db 321 ASPDFLPGSOHKVKVWIDSTKLAL 345

RESULT 15
Q9SYF4 PRELIMINARY; PRT; 91 AA.
AC Q9SYF4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE F1511.6 PROTEIN.
GN F1511.6.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Li J., Kremenetskaia I., Luros J., Ngan I., Gonzalez A., Altafi H.,
RA Araujo R., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,
RA Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC F1511 sequence.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC006577; AAD25769.1; -.
DR InterPro: IPR000626; -.
DR Pfam: PF00240; ubiqtutin; 2.
DR SEQUENCE 91 AA; 10142 MW; E8766823D6450267 CRC64;

Query Match 41.5%; Score 44; DB 10; Length 91;
Best Local Similarity 47.6%; Pred. No. 5.8;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

```



```

RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RL region from map positions 64% to 92% of the genome." ;
RN DNA Res. 2:153-166(1995).
RX [3]
RY SEQUENCE FROM N.A.
RZ MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RM Miyajima N., Hirose M., Suglura M., Sasamoto S., Kimura T.,
RS Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpō S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RT Tabata S.;
RV "Sequence analysis of the genome of the unicellular cyanobacterium
RW Synechocystis sp. strain PCC6803. II. Sequence determination of the
RX entire genome and assignment of potential protein-coding regions.";
RY DNA Res. 3:109-136(1996).
RZ EMBL; D63999; BAA10088.1; -.
RA InterPro; IPR002048; -.
RM PROSITE; PS00018; EF_HAND; UNKNOWN_4.
RZ Hypothetical protein.
RY KW
RZ SEQUENCE 1742 AA; 179710 MW; 56C815A887D30B82 CRC64;
SY QUERY MATCH 41.5%; SCORE 44; DB 2; LENGTH 1742;
BEST LOCAL SIMILARITY 38.1%; PRED. NO. 1.5e+02;
MATCHES 8; CONSERVATIVE 4; MISMATCHES 9; INDELS 0; GAPS 0;

QY 1 GATFQVEVPQSQHIDSQKKAI 21
DB 121 GGTFSIDVPGSALVDPDRTI 141
III::IIII::I

RESULT 20
ID Q22511 PRELIMINARY; PRT; 565 AA.
AC Q22511;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JUN-1998 (TREMBlrel. 05, Last sequence update)
DE 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE GLUTATHIONE REDUCTASE (NADPH) (EC 1.6.4.2) (FRAGMENT).
OS GOR.
OS Vitis vinifera (Grape).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OX Magnoliophyta; eudicotyledons; core eudicots; Vitaceae; Vitis.
OX NCBI_TaxId=29760;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. THOMPSON SEEDLESS (CLONE 2A); TISSUE=FRUIT;
RA Cassol T., Adams D.O.;
RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
CC -I- COFACTOR: FAD (BY SIMILARITY).
CC -I- SIMILARITY: TO PYRIDINE NUCLEOTIDE-DISULPHIDE OXIDOREDUCTASES
CC CLASS-I.
DR EMBL; AF019907; AAB70837.1; -.
DR HSP; P00390; IALG.
DR Mendel; 26381; Vitv1;190.26381.
DR InterPro; IPR000103; -.
DR InterPro; IPR001100; -.
DR InterPro; IPR001327; -.
DR Pfam; PF00070; Pyr_redox; 1.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00411; PNDRITASEI.
DR PRINTS; PR00469; PNDRITASEII.
DR PROSITE; PS00076; PYRIDINE_REDOX_1; 1.
KW FAD; Flavoprotein; Oxidoreductase; Redox-active center.
FT NON_TER 1
SQ SEQUENCE 565 AA; 60695 MW; B26113AE09A121DE CRC64;

QUERY MATCH 41.0%; SCORE 43.5; DB 10; LENGTH 565;
BEST LOCAL SIMILARITY 58.8%; PRED. NO. 52;
MATCHES 10; CONSERVATIVE 2; MISMATCHES 4; INDELS 1; GAPS 1;

QY 1 GATFQVEVPQSQH-IDS 16
```

```
DR InterPro; IPR002123; -.
DR Pfam; PF01553; Acyltransferase; 1.
KW Transferase.
SQ SEQUENCE 356 AA; 39755 MW; FE5C9A120287F299 CRC64;

Query Match 40.6%; Score 43; DB 5; Length 356;
Best Local Similarity 38.9%; Pred. No. 38;
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 4 FQVEVPGSQHIDSQKAI 21
| : | : | : | : | : |
Db 186 FKSDADGNFHDVDEKQAI 203

RESULT 23
Q27066 PRELIMINARY; PRT; 365 AA.
ID O27066
AC O27066;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE SENSORY TRANSDUCTION HISTIDINE KINASE.
GN MTH985.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucet-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,
RA Alredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwan N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noll J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT delta: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000872; AAB85482.1; -.
DR InterPro; IPR000014; -.
DR InterPro; IPR000410; -.
DR InterPro; IPR000700; -.
DR InterPro; IPR003594; -.
DR Pfam; PF00512; signal; 1.
DR Pfam; PF00989; PAS; 1.
DR SMART; SM00387; HATPase_c; 1.
SQ SEQUENCE 365 AA; 41615 MW; 50C56FE519250714 CRC64;

Query Match 40.6%; Score 43; DB 1; Length 365;
Best Local Similarity 36.8%; Pred. No. 39;
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Qy 2 ATFQVEVPGSQHIDSQKAI 20
| : | : | : | : | : |
Db 49 AIFLVKIPGGELVDANRSA 67

RESULT 24
Q9F5L7 PRELIMINARY; PRT; 399 AA.
ID Q9F5L7
AC Q9F5L7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE TRNA GUANINE TRANSGLYCOSYLASE.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;
OC Zymomonas.

OX NCBI_TaxID=542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ZM4;
RA Ahn J.Y., Kang H.S.;
RT "Sequence analysis of 44B6 fosmid clone of Zymomonas mobilis ZM4.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF313764; AAG29862.1; -.
SQ SEQUENCE 399 AA; 44287 MW; 75CEF229E70F94A7 CRC64;

Query Match 40.6%; Score 43; DB 2; Length 399;
Best Local Similarity 38.1%; Pred. No. 43;
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 1 GATFQVEVPGSQHIDSQKAI 21
| : | : | : | : | : |
Db 134 GVTFKSHLDGSRHMLSPERSI 154

RESULT 25
Q9LY67 PRELIMINARY; PRT; 439 AA.
ID Q9LY67
AC Q9LY67;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE KINESIN HEAVY CHAIN-LIKE PROTEIN.
GN MAZ1_110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC Rieger M., Mueller-Auer S., Zipp M., Schaefer M., Mewes H.W., Rudd S.,
RA Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL163818; CAB87801.1; -.
DR InterPro; IPR001752; -.
DR Pfam; PF00225; kinesin; 2.
DR PRINTS; PR00380; KINESINHEAVY.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN2; 1.
DR SMART; SM00129; KISC; 1.
KW ATP-binding; Coiled coil; Microtubules; Motor protein.
SQ SEQUENCE 439 AA; 49028 MW; 9BD55085A1966D70 CRC64;

Query Match 40.6%; Score 43; DB 10; Length 439;
Best Local Similarity 38.1%; Pred. No. 48;
Matches 8; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

Qy 1 GATFQVEVPGSQHIDSQKAI 21
| : | : | : | : | : |
Db 91 GKTYSMEGPGIQDCDEHNKGL 111

Search completed: July 16, 2001, 16:43:38
Job time: 472 sec
```



***This Page Blank (uspto)***



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 16:35:41 ; Search time 56.06 Seconds  
(without alignments)  
22.710 Million cell updates/sec

Title: US-09-786-648-5

Perfect score: 107

Sequence: 1 GETQVEVPGSQHDSQKAI 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*
- 22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	21	AA1987463	E. coli heat labil
2	107	100.0	93	AA19872545	ADP-ribosylating t
3	107	100.0	93	AA19871816	Escherichia coli v
4	107	100.0	93	AA19875226	E. coli heat-labil
5	107	100.0	93	AA19873655	Heat labile toxin
6	107	100.0	93	AA19876239	E coli verotoxin-1
7	107	100.0	134	AA19873241	Recombinant exotox
8	107	100.0	142	AA19873242	Recombinant exotox
9	107	100.0	155	AA19873243	Recombinant exotox
10	107	100.0	163	AA19873244	Recombinant exotox
11	107	100.0	371	AA19871300	Labile toxin (LT-B

12	107	100.0	371	20	AA1987443	C. jejuni flagelli
13	101	94.4	21	21	AA1987462	Cholera toxin B/en
14	101	94.4	103	17	AA1987457	Synthetic cholera
15	101	94.4	103	17	AA1987457	Heat labile entero
16	101	94.4	103	17	AA1987457	Cholera toxin B su
17	101	94.4	103	17	AA1987457	Cholera toxin B su
18	101	94.4	103	19	AA1987457	Amino acid sequenc
19	101	94.4	118	11	AA1987457	Cholera Toxin B-su
20	101	94.4	124	10	AA1987457	B subunit of the h
21	101	94.4	124	17	AA1987457	Cholera toxin B su
22	101	94.4	124	19	AA1987457	Amino acid sequenc
23	101	94.4	124	21	AA1987457	Plant-optimized E.
24	101	94.4	131	11	AA1987457	LTB-CTP fusion pro
25	101	94.4	138	15	AA1987457	Sequence of LT-B-M
26	101	94.4	170	20	AA1987457	LTB-CTP fusion pro
27	101	94.4	405	12	AA1987457	HSV-1 antigen/heat
28	98	91.6	103	6	AA1987457	Sequence of sub-un
29	95	88.8	124	21	AA1987457	Plant-optimized V.
30	95	88.8	126	12	AA1987457	GtFA.1/CTB chimeri
31	94	87.9	41	6	AA1987457	Network polymer wh
32	89	83.2	124	13	AA1987457	B subunit of CT.
33	89	83.2	461	19	AA1987457	Adhesin/V.cholerae
34	89	83.2	749	19	AA1987457	Helicobacter pylor
35	89	83.2	1338	19	AA1987457	Helicobacter pylor
36	88	82.2	46	6	AA1987457	Network polymer wh
37	87	81.8	47	4	AA1987457	Sequence of amino
38	82	76.6	461	20	AA1987457	Adhesin/CTXA2B chi
39	80	74.8	26	4	AA1987457	Sequence of amino
40	76	71.0	15	10	AA1987457	CTP3 epitope of th
41	76	71.0	15	16	AA1987457	Cholera toxin B an
42	76	71.0	23	16	AA1987457	Residues 50-64 of
43	62	57.9	12	21	AA1987457	Cholera toxin B/en
44	61	57.0	15	21	AA1987457	Cholera toxin B su
45	61	57.0	15	21	AA1987457	Bovine rotavirus V

## ALIGNMENTS

### RESULT 1

AA1987463

ID AA1987463 standard; peptide; 21 AA.

AC AA1987463;

XX

DT 03-JUL-2000 (first entry)

XX

DE E. coli heat labile enterotoxin B-derived peptide, SEQ ID NO:5.

XX

KW Heat labile enterotoxin subunit B; EtXB;

KW beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;

XX adjuvant; immune disorder; diarrhoea.

XX Escherichia coli.

XX WO200014114-A1.

PN

XX

PD 16-MAR-2000.

XX

PF 07-SEP-1999; 99WO-GB02970.

XX

PR 07-SEP-1998; 98GB-0019484.

XX

PA (UYBR-) UNIV BRISTOL.

XX

PI Williams NA, Hirst TR;

XX

DR WPI; 2000-256943/22.

XX

PT Derivatives of Escherichia coli heat labile enterotoxins useful as immunomodulators and for treating diarrhea and which do not bind the glycolipid receptor GM-1 -

XX

PS Disclosure; Page 15; 62pp; English.

XX The invention relates to peptide fragments of the *Escherichia coli* heat  
CC labile enterotoxin (Etx) and its closely related homologue, cholera  
CC toxin (Ctx) from *Vibrio cholerae* which do not bind to the ubiquitous  
CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are  
CC composed of one A subunit and five identical B subunits. The A subunit  
CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-  
CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)  
CC facilitate the entry of subunit A into the host cell via the binding and  
CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible  
CC for some of the effects of Etx and Ctx, it has been found that certain  
CC effects of the toxins, such as immunomodulation, are not mediated  
CC through GM-1 binding. The peptides of the invention are fragments of the  
CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as  
CC normal EtxB and CtxB subunits, except that they do not bind or cross link  
CC GM-1. They may be used as an inhibitor for toxin-induced diarrhoea.  
CC They may also be used as an inhibitor for immunomodulation or adjuvant.  
CC Therefore, the peptides may be used in the production of a composition  
CC for treating, preventing and/or modulating a disease associated with an  
CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463  
CC represent preferred peptides of the invention, AAY87460 being  
CC particularly preferred.

XX Sequence 21 AA;

Query Match 100.0%; Score 107; DB 21; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFQVEVPGSQHDSQKKAI 21  
Db 1 getfqvpgsqhdsqkai 21  
|||||

RESULT 2

AA72545  
ID AAR72545 standard; peptide; 93 AA.

AC AAR72545;

DT 28-NOV-1995 (first entry)

DE ADP-ribosylating toxin (verotoxin-1 B-subunit).

KW ADP-ribosylating toxin; pertussis holotoxin; B-subunit;

KW active site; *E. coli* heat labile toxin; verotoxin-1;

KW Bordetella pertussis vaccines.

OS Bacteria sp.

PN EP646599-A.

XX 05-APR-1995.

PF 23-AUG-1994; 94EP-0306219.

PR 24-AUG-1993; 93US-0110947.

PR 31-MAY-1994; 94US-0251121.

XX (CONN-) CONNAUGHT LAB LTD.

PA (UYAL-) UNIV ALBERTA.

XX Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;

PI Oomen R, Read RJ, Stein PE;

XX WPI; 1995-132623/18.

XX New modified forms of pertussis holotoxin - developed using  
PT crystalline forms of pertussis holotoxin and its complexes with  
PT other molecules

PS Disclosure; Fig 5; 54pp; English.

XX AAR72540-872545 are structurally equivalent B-subunits from three  
CC ADP-ribosylating toxins, pertussis holotoxin (PT), *E. coli* heat  
CC labile toxin (LT), and verotoxin-1 (VT). The structural  
CC information obtained from these comparisons was used to identify  
CC sites which contribute to PT's biological activity. By modifying  
CC these sites the claimed PT mutants of the invention were produced,  
CC they can be used in the development of vaccines against Bordetella  
CC pertussis infection.

XX Sequence 93 AA;

Query Match 100.0%; Score 107; DB 16; Length 93;

Best Local Similarity 100.0%; Pred. No. 8.1e-11;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFQVEVPGSQHDSQKKAI 21

Db 35 getfqvpgsqhdsqkai 55

|||||

RESULT 3

AA72545

ID AAY41816 standard; peptide; 93 AA.

XX AAY41816;

DT 08-DEC-1999 (first entry)

DE *Escherichia coli* verotoxin-1 B-subunit.

KW ADP-ribosylating toxin; PT; crystalline pertussis holotoxin; ETA; DT;  
KW three-dimensional structure; LT; immunoprotective; infection.

OS *Escherichia coli*.

PN US5965385-A.

XX 12-OCT-1999.

XX 06-JUN-1995; 95US-0467974.

XX 22-AUG-1994; 94US-0292968.

XX 24-AUG-1993; 93US-0110947.

XX 31-MAY-1994; 94US-0251121.

XX (CONN-) CONNAUGHT LAB LTD.

PA (UYAL-) UNIV ALBERTA.

XX Read RJ, Stein PE, Cockle SA, Loosmore S, Klein MH, Armstrong GD;

PI Hazes B, Oomen RP;

XX WPI; 1999-579908/49.

XX New method for producing modified pertussis holotoxin -

XX Example 3; Fig 5; 41pp; English.

XX A method has been developed of producing a modified pertussis holotoxin,  
CC involving analysis of the 3-dimensional form of the crystalline  
CC holotoxin. The pertussis holotoxin modification process comprises:

CC (1) identification of at least one amino acid (aa) residue of the  
CC holotoxin for modification by analysing the 3-dimensional form of the  
CC crystalline holotoxin, in relation to known information of the protein  
CC structure and function; (2) effecting mutagenesis (by removing or  
CC replacing a nucleotide sequence encoding at least one (aa) of a tox  
CC operon; and (3) expressing mutant toxin in a Bordetella organism to  
CC produce the modified holotoxin. This method is used for modifying  
CC pertussis holotoxin, by studying its 3-dimensional crystalline  
CC structure. Modifying the holotoxin, alters its biological properties.  
CC By analysing the 3-dimensional crystalline structure of the pertussis

CC holotoxin, functional (aa) which affect biological properties of the  
 CC pertussis holotoxin can be identified. This can be used to predict (aa)  
 CC which contribute to the toxicity of the holotoxin to produce  
 CC immunoprotective, genetically-detoxified analogues of pertussis  
 CC holotoxin. The present sequence represents an ADP-ribosylating toxin  
 CC B-subunit peptide used in the exemplification of the present  
 CC invention.  
 CC Sequence 93 AA;  
 SQ

Query Match 100.0%; Score 107; DB 20; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 8.1e-11;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFQVEVPGSQHDSQKRAI 21  
 Db 35 getfvevpgsqhidsqkai 55  
 |||||

RESULT 4  
 AAW95226  
 ID AAW95226 standard; peptide; 93 AA.  
 XX AC AAW95226;  
 XX DT 16-MAR-1999 (first entry)  
 XX DE E. coli heat-labile toxin (LT) beta-subunit sequence.  
 XX KW Pertussis holotoxin; PT; modified; effector; toxicity; cell binding;  
 KW enzyme; ADP-ribosylation; toxin; crystalline; X-ray crystallography;  
 KW structural analysis; interacting site; mitogenicity; adjuvant activity;  
 KW heat-labile; LT.  
 XX OS Escherichia coli.  
 XX PN US5856122-A.  
 XX PD 05-JAN-1999.  
 XX PF 22-AUG-1994; 94US-0292968.  
 XX PR 22-AUG-1994; 94US-0292968.  
 XX PR 24-AUG-1993; 93US-0110947.  
 XX PR 31-MAY-1994; 94US-0251121.  
 XX PA (UYAL-) UNIV ALBERTA.  
 XX PI Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;  
 PI Oomen RP, Read RJ, Stein PE;  
 XX WPI; 1999-105104/09.  
 XX Modifications to e.g. enzymatic activity, mitogenicity and cell  
 PT binding of pertussis holotoxin - by identifying interaction sites of  
 PT a molecule with crystalline toxin and modifying the identified site  
 XX Example 3; Fig 5; 40pp; English.  
 XX The invention relates to methods of preparing a pertussis holotoxin (PT)  
 CC having a modified biological activity. One method comprises identifying  
 CC at least 1 site in a PT that interacts with a molecule that is capable of  
 CC forming a complex with the holotoxin and which molecule is an effector  
 CC molecule which is an adenine nucleotide and which site contributes to  
 CC toxicity, cell binding or enzymatic activity of PT. The functional  
 CC interacting site(s) are identified by analysing the three dimensional  
 CC structure of crystalline PT, determined by X-ray crystallography. The  
 CC identified interacting site(s) are modified to alter toxicity, cell  
 CC binding or enzyme activity of the PT. The methods can be used to alter a  
 CC biological activity such as toxicity, enzymatic activity, mitogenicity,  
 CC cell binding and adjuvant activity of the PT. The three-dimensional structure  
 CC of PT have functional and/or structural resemblance to other bacterial

CC toxins such as diphtheria toxin (DT), Pseudomonas exotoxin A (ETA), the  
 CC heat-labile toxin of E. coli (LT) and verotoxin-1 (VT). The present  
 CC sequence represents the beta-subunit of LT toxin.  
 XX Sequence 93 AA;  
 SQ

Query Match 100.0%; Score 107; DB 20; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 8.1e-11;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFQVEVPGSQHDSQKRAI 21  
 Db 35 getfvevpgsqhidsqkai 55  
 |||||

RESULT 5  
 AAY68365  
 ID AAY68365 standard; Peptide; 93 AA.  
 XX AC AAY68365;  
 XX DT 17-APR-2000 (first entry)  
 XX DE Heat labile toxin B subunit SEQ ID NO:26.  
 XX KW Pertussis toxin; heat labile toxin; verotoxin; exotoxin A; holotoxin;  
 KW diphtheria toxin; ADP-ribosylating toxin; mannose binding protein;  
 KW infection; crystal structure; X-ray crystallography; detoxification;  
 KW immunogenic.  
 XX OS Escherichia coli.  
 XX PN US6018022-A.  
 XX PD 25-JAN-2000.  
 XX PF 06-JUN-1995; 95US-0467976.  
 XX PR 22-AUG-1994; 94US-0292968.  
 XX PR 24-AUG-1993; 93US-0110947.  
 XX PR 31-MAY-1994; 94US-0251121.  
 XX PA (CONN-) CONNAUGHT LAB LTD.  
 XX PA (UYAL-) UNIV ALBERTA.  
 XX PI Read RJ, Stein PE, Cockle SA, Loosmore S, Klein MH, Armstrong GD;  
 PI Hazes B, Oomen RP;  
 XX WPI; 2000-136703/12.  
 XX Crystalline form of isolated pertussis holotoxin useful in studying  
 PT proteins which have functional resemblance -  
 XX Example 3; Fig 5; 42pp; English.  
 XX The present invention describes a crystalline form of isolated  
 CC pertussis holotoxin, in which the molecules of pertussis toxin have  
 CC a three dimensional structure represented in the specification,  
 CC complexed with a polysaccharide molecule capable of forming a complex  
 CC with the holotoxin. The crystalline form of the pertussis holotoxin  
 CC can be used in a comparison with other proteins which have functional  
 CC resemblance to pertussis holotoxin with the aim of modifying other  
 CC proteins. Identifying the unknown sites of toxicity by comparison  
 CC with the three dimensional structure of pertussis holotoxin provides a  
 CC technique for detoxification of toxins to produce useful immunogenic  
 CC but non-toxic analogues. It can also be used as a primary standard to  
 CC measure the quantity, purity or efficacy of less pure compositions  
 CC containing pertussis toxin. AAY68340 to AAY68385 represent peptides  
 CC used in the exemplification of the present invention.  
 XX Sequence 93 AA;  
 SQ

Query Match 100.0%; Score 107; DB 21; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 8.1e-11;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSQHDSQKKAI 21  
 Db 35 getfgvevpgsqhdsqkkai 55

## RESULT 6

AAB66239  
 ID AAB66239 standard; Protein; 93 AA.

XX AC AAB66239;

XX DT 03-APR-2001 (first entry)

XX DE E coli verotoxin-1 B subunit SEQ ID NO: 26.

XX KW Pertussis toxin; crystal structure; whooping cough; biological activity;  
 KW lymphocytosis-promoting factor; histamine-sensitising factor;  
 KW islet-activating protein.

XX OS Escherichia coli.

XX PN US6168928-B1.

XX PD 02-JAN-2001.

XX PF 21-MAY-1998; 98US-0082514.

XX PR 22-AUG-1994; 94US-0292968.

XX PR 24-AUG-1993; 93US-0110947.

XX PR 31-MAY-1994; 94US-0251121.

XX PA (CONN-) CONNAUGHT LAB LTD.

XX PI Read RJ, Cockle SA, Oomen RP, Loosmore S, Klein MH, Armstrong GD;  
 PI Hazes B, Stein PE;

XX DR WPI; 2001-122260/13.

XX PT Modifying pertussis holotoxin to produce detoxified PT analogs,  
 PT comprising analyzing crystalline structure of toxin, to identify sites  
 PT of toxicity, cell binding or enzyme activity of PT and modifying  
 PT identified site.

XX PS Example 3; Fig 5; 41pp; English.

XX CC The present invention provides a method for producing a pertussis toxin  
 CC (also designated lymphocytosis-promoting factor, histamine-sensitising  
 CC factor and islet activating protein) with a modified biological activity,  
 CC involving analysing the crystal structure of the protein to identify  
 CC active sites which can then be modified. This may lead to an alteration  
 CC in the toxicity, cell binding or enzyme activity of the toxin. This can  
 CC be used in the production of immunoprotective analogues of pertussis  
 CC toxin. Pertussis toxin is the cause of whooping cough following infection  
 CC by Bordetella pertussis.

XX SQ Sequence 93 AA;

Query Match 100.0%; Score 107; DB 22; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 8.1e-11;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSQHDSQKKAI 21  
 Db 35 getfgvevpgsqhdsqkkai 55

## RESULT 7

## AAB73241

ID AAB73241 standard; Protein; 134 AA.

XX AC AAB73241;

XX DT 14-MAY-2001 (first entry)

XX DE Recombinant exotoxin protein variant LTbPL.

XX KW Exotoxin mucosal cell binding motif; nucleic acid delivery;  
 KW nucleic acid affinity domain; heat-labile enterotoxin.

XX OS Unidentified.

XX PN WO200111960-A1.

XX PD 22-FEB-2001.

XX PF 18-AUG-2000; 2000WO-US22715.

XX PR 18-AUG-1999; 99US-0149294.

XX PA (AGRI-) AGRIVAX INC.

XX PI Weiter LM;

XX DR WPI; 2001-211103/21.

XX DR N-PSDB; AAF75712.

XX PT Novel exotoxin protein variant useful as protein carrier for  
 PT facilitating gene delivery, comprises a mucosal cell binding motif of  
 PT an exotoxin and a nucleic acid affinity domain.

XX PS Example 1; Fig 9; 57pp; English.

XX CC The present invention relates to recombinant exotoxin protein variants,  
 CC which comprise an exotoxin mucosal cell binding motif and a nucleic acid  
 CC affinity domain. The present sequence is one such protein variant. In the  
 CC present invention the heat-labile enterotoxin (LT) of Escherichia coli  
 CC was used as an exotoxin. LT is a member of the ADP-ribosylating exotoxin  
 CC (BARE) family. The protein variants are useful for selectively delivering  
 CC nucleic acid to mucosal cells, for inducing an immune response when the  
 CC nucleic acid encodes an antigen to which the immune response is desired,  
 CC for selectively delivering a gene to a mucosal cell, and for achieving  
 CC expression of a protein in a subject, by administering a composition  
 CC comprising the protein variant.

XX SQ Sequence 134 AA;

Query Match 100.0%; Score 107; DB 22; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSQHDSQKKAI 21  
 Db 66 getfgvevpgsqhdsqkkai 86

## RESULT 8

AAB73242

ID AAB73242 standard; Protein; 142 AA.

XX AC AAB73242;

XX DT 14-MAY-2001 (first entry)

XX DE Recombinant exotoxin protein variant LTbPL.

XX KW Exotoxin mucosal cell binding motif; nucleic acid delivery;  
 KW nucleic acid affinity domain; heat-labile enterotoxin.

XX OS Unidentified.

XX WO200111960-A1.  
PN  
XX  
XX 22-FEB-2001.  
PD  
XX  
XX 18-AUG-2000; 2000WO-US22715.  
PF  
XX  
XX 18-AUG-1999; 99US-0149294.  
PR  
XX  
XX (AGRI-) AGRIVAX INC.  
PA  
XX  
XX Welter LM;  
PI  
XX  
XX WPI: 2001-211103/21.  
DR  
XX N-PSDB; AAF75713.  
XX  
XX Novel exotoxin protein variant useful as protein carrier for  
PT facilitating gene delivery, comprises a mucosal cell binding motif of  
PT an exotoxin and a nucleic acid affinity domain  
XX  
XX Example 1; Fig 10; 57pp; English.  
PS  
XX The present invention relates to recombinant exotoxin protein variants,  
CC which comprise an exotoxin mucosal cell binding motif and a nucleic acid  
CC affinity domain. The present sequence is one such protein variant. In the  
CC present invention the heat-labile enterotoxin (LT) of Escherichia coli  
CC was used as an exotoxin. LT is a member of the ADP-ribosylating exotoxin  
CC (bARE) family. The protein variants are useful for selectively delivering  
CC nucleic acid to mucosal cells, for inducing an immune response when the  
CC nucleic acid encodes an antigen to which the immune response is desired,  
CC for selectively delivering a gene to a mucosal cell, and for achieving  
CC expression of a protein in a subject, by administering a composition  
CC comprising the protein variant.  
XX  
XX Sequence 142 AA;  
SQ

Query Match 100.0%; Score 107; DB 22; Length 142;  
Best Local Similarity 100.0%; Pred. No. 1.3e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFQVEVPGSQHDSQKRAI 21  
Db 66 getfqvpgsqhidsqgkai 86  
|||||

RESULT 9  
AAB73243  
ID AAB73243 standard; Protein; 155 AA.  
XX  
XX AAB73243;  
AC  
XX  
XX 14-MAY-2001 (first entry)  
DT  
XX Recombinant exotoxin protein variant LTB-P.  
DE  
XX Exotoxin mucosal cell binding motif; nucleic acid delivery;  
KW nucleic acid affinity domain; heat-labile enterotoxin.  
XX  
XX Unidentified.  
OS  
XX WO200111960-A1.  
PN  
XX  
XX 22-FEB-2001.  
PD  
XX  
XX 18-AUG-2000; 2000WO-US22715.  
PF  
XX  
XX 18-AUG-1999; 99US-0149294.  
PR  
XX  
XX (AGRI-) AGRIVAX INC.  
PA  
XX  
XX Welter LM;  
PI  
XX  
XX WPI: 2001-211103/21.  
DR  
XX N-PSDB; AAF75713.  
XX  
XX Novel exotoxin protein variant useful as protein carrier for  
PT facilitating gene delivery, comprises a mucosal cell binding motif of  
PT an exotoxin and a nucleic acid affinity domain  
XX  
XX Example 1; Fig 12; 57pp; English.  
PS  
XX The present invention relates to recombinant exotoxin protein variants,  
CC which comprise an exotoxin mucosal cell binding motif and a nucleic acid  
CC affinity domain. The present sequence is one such protein variant. In the  
CC present invention the heat-labile enterotoxin (LT) of Escherichia coli  
CC was used as an exotoxin. LT is a member of the ADP-ribosylating exotoxin  
CC (bARE) family. The protein variants are useful for selectively delivering  
CC nucleic acid to mucosal cells, for inducing an immune response when the  
CC nucleic acid encodes an antigen to which the immune response is desired,  
CC for selectively delivering a gene to a mucosal cell, and for achieving  
CC expression of a protein in a subject, by administering a composition  
CC comprising the protein variant.  
XX  
XX Sequence 142 AA;  
SQ

DR WPI: 2001-211103/21.  
DR N-PSDB; AAF75714.  
XX  
XX Novel exotoxin protein variant useful as protein carrier for  
PT facilitating gene delivery, comprises a mucosal cell binding motif of  
PT an exotoxin and a nucleic acid affinity domain  
XX  
XX Example 1; Fig 11; 57pp; English.  
PS  
XX The present invention relates to recombinant exotoxin protein variants,  
CC which comprise an exotoxin mucosal cell binding motif and a nucleic acid  
CC affinity domain. The present sequence is one such protein variant. In the  
CC present invention the heat-labile enterotoxin (LT) of Escherichia coli  
CC was used as an exotoxin. LT is a member of the ADP-ribosylating exotoxin  
CC (bARE) family. The protein variants are useful for selectively delivering  
CC nucleic acid to mucosal cells, for inducing an immune response when the  
CC nucleic acid encodes an antigen to which the immune response is desired,  
CC for selectively delivering a gene to a mucosal cell, and for achieving  
CC expression of a protein in a subject, by administering a composition  
CC comprising the protein variant.  
XX  
XX Sequence 155 AA;  
SQ

Query Match 100.0%; Score 107; DB 22; Length 155;  
Best Local Similarity 100.0%; Pred. No. 1.5e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFQVEVPGSQHDSQKRAI 21  
Db 66 getfqvpgsqhidsqgkai 86  
|||||

RESULT 10  
AAB73244  
ID AAB73244 standard; Protein; 163 AA.  
XX  
XX AAB73244;  
AC  
XX  
XX 14-MAY-2001 (first entry)  
DT  
XX Recombinant exotoxin protein variant LTB-Ph.  
DE  
XX Exotoxin mucosal cell binding motif; nucleic acid delivery;  
KW nucleic acid affinity domain; heat-labile enterotoxin; ds.  
XX  
XX Unidentified.  
OS  
XX WO200111960-A1.  
PN  
XX  
XX 22-FEB-2001.  
PD  
XX  
XX 18-AUG-2000; 2000WO-US22715.  
PF  
XX  
XX 18-AUG-1999; 99US-0149294.  
PR  
XX  
XX (AGRI-) AGRIVAX INC.  
PA  
XX  
XX Welter LM;  
PI  
XX  
XX WPI: 2001-211103/21.  
DR  
XX N-PSDB; AAF75715.  
XX  
XX Novel exotoxin protein variant useful as protein carrier for  
PT facilitating gene delivery, comprises a mucosal cell binding motif of  
PT an exotoxin and a nucleic acid affinity domain  
XX  
XX Example 1; Fig 12; 57pp; English.  
PS  
XX The present invention relates to recombinant exotoxin protein variants,  
CC which comprise an exotoxin mucosal cell binding motif and a nucleic acid  
CC affinity domain. The present sequence is one such protein variant. In the  
CC present invention the heat-labile enterotoxin (LT) of Escherichia coli  
CC

CC was used as an exotoxin. LT is a member of the ADP-ribosylating exotoxin  
 CC (bare) family. The protein variants are useful for selectively delivering  
 CC nucleic acid to mucosal cells, for inducing an immune response when the  
 CC nucleic acid encodes an antigen to which the immune response is desired,  
 CC for selectively delivering a gene to a mucosal cell, and for achieving  
 CC expression of a protein in a subject, by administering a composition  
 CC comprising the protein variant.  
 XX  
 SQ Sequence 163 AA;

Query Match 100.0%; Score 107; DB 22; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSOHIDSOKKAI 21  
 |||||  
 DB 66 getfqvevpgsqhidsqkai 86

RESULT 11  
 AAV01300  
 ID AAV01300 standard; Protein; 371 AA.

XX  
 AC AAV01300;

DT 03-JUN-1999 (first entry)

DE Labile toxin (LT-B)/flagellin (flaA) fusion protein.

XX Fusion protein; labile toxin B subunit; LT-B; E. coli; flagellin; flaA;  
 KW C. jejuni; antigenic; colonisation; chicken; Campylobacter; vaccine;  
 KW immune response; ganglioside GM1; immunoglobulin A.

OS Campylobacter jejuni.  
 OS Escherichia coli.

XX US5888810-A.

PN 30-MAR-1999.

XX PF 16-JAN-1997; 97US-0784218.

XX PR 12-NOV-1993; 93US-0150305.

XX PR 16-JAN-1997; 97US-0784218.

XX PA (USDA ) US SEC OF AGRIC.

XX PI Khoury CA, Meinersmann RJ;

XX WPI; 1999-243214/20.

XX N-PSDB; AAX27771.

XX PT Recombinant fusion gene expressing Campylobacter jejuni flagellin  
 fragment

XX PS Disclosure; Column 16-20; 14pp; English.

XX This represents a fusion protein that comprises the B subunit of labile  
 CC toxin (LT-B) of E. coli and part of the flagellin (flaA) protein of  
 CC C. jejuni. The fusion protein is antigenic and is useful for decreasing  
 CC colonisation in chickens by Campylobacter species. The labile toxin  
 CC B-subunit component of the fusion protein (i) binds to ganglioside GM1  
 CC on the surface of eukaryotic cells, delivering the antigenic flagellin  
 CC component to the mucosal surface, resulting in an immune response against  
 CC Campylobacter jejuni reducing colonization by the organism; and (ii) has  
 CC an adjuvant effect on immunoglobulin A secretion. The fusion protein  
 CC expressed by the recombinant gene is relatively simple to produce and  
 CC purify, can be produced in large quantities and can be used for  
 CC vaccination without any further treatment other than purification.  
 XX  
 SQ Sequence 371 AA;

Query Match 100.0%; Score 107; DB 20; Length 371;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSOHIDSOKKAI 21  
 |||||  
 DB 49 getfqvevpgsqhidsqkai 69

RESULT 12  
 AAW67443  
 ID AAW67443 standard; Protein; 371 AA.

XX AC AAW67443;

DT 02-MAR-1999 (first entry)

DE C. jejuni flagellin/E. coli LT-B fusion protein.

XX Recombinant; fusion protein; flagellin; flaA; labile toxin; antigen;  
 KW vaccine; colonisation; chicken; mucosal surface; adjuvant; secretion;  
 KW immunoglobulin A.

XX Chimeric - Campylobacter jejuni.

OS Chimeric - Escherichia coli.

XX US5837825-A.

XX PD 17-NOV-1998.

XX PF 31-MAR-1997; 97US-0829026.

XX PR 12-NOV-1993; 93US-0150305.

XX PR 31-MAR-1997; 97US-0829026.

XX PA (USDA ) US SEC OF AGRIC.

XX PI Khoury CA, Meinersmann RJ;

XX WPI; 1999-023536/02.

XX N-PSDB; AAV34380.

XX PT Recombinant fusion protein of Campylobacter jejuni flagellin protein  
 and Escherichia coli labile toxin - useful in vaccines to reduce  
 PT colonisation of chickens by Campylobacter species

XX PS Disclosure; Column 17-20; 14pp; English.

XX This sequence represents a recombinant fusion protein consisting of the  
 CC Campylobacter jejuni flagellin protein (flaA) fused to the Escherichia  
 CC coli B subunit of the labile toxin (LT-B). The fusion protein is  
 CC antigenic and when administered in vaccines decreases colonisation of  
 CC chickens by Campylobacter species. The LT-B component serves to deliver  
 CC the flagellin antigen to mucosal surfaces and also has an adjuvant effect  
 CC on immunoglobulin A secretion.

XX SQ Sequence 371 AA;

Query Match 100.0%; Score 107; DB 20; Length 371;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSOHIDSOKKAI 21  
 |||||  
 DB 49 getfqvevpgsqhidsqkai 69

RESULT 13  
 AAY87462  
 ID AAY87462 standard; peptide; 21 AA.  
 XX

AC	RAY87462;
XX	
DT	03-JUL-2000 (first entry)
XX	
DE	Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:4.
XX	
KW	Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;
KW	beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;
KW	adjuvant; immune disorder; diarrhoea.
XX	
OS	Vibrio cholerae.
OS	Escherichia coli.
XX	
PN	WO200014114-A1.
XX	
PD	16-MAR-2000.
XX	
PY	07-SEP-1999; 99WO-GS02970.
PF	
PR	07-SEP-1998; 98GB-0019484.
XX	
PA	(UYBR-) UNIV BRISTOL.
XX	
PI	Williams NA, Hirst TR;
XX	
PS	WPI; 2000-256943/22.
XX	
PT	Derivatives of Escherichia coli heat labile enterotoxins useful as
PT	immunomodulators and for treating diarrhoea and which do not bind the
PT	glycolipid receptor GM-1 .
XX	
PS	Disclosure; Page 15; 62pp; English.
XX	
CC	The invention relates to peptide fragments of the Escherichia coli heat
CC	labile enterotoxin (Etx) and its closely related homologue, cholera
CC	toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous
CC	GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are
CC	composed of one A subunit and five identical B subunits. The A subunit
CC	is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-
CC	ribosyltransferase activity, while the B subunits (ETxB and CTxB)
CC	facilitate the entry of subunit A into the host cell via the binding and
CC	cross-linking of GM-1 receptors. Although GM-1 binding is responsible
CC	for some of the effects of Etx and Ctx, it has been found that certain
CC	effects of the toxins, such as immunomodulation, are not mediated
CC	through GM-1 binding. The peptides of the invention are fragments of the
CC	beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as
CC	normal EtxB and CtxB subunits, except that they do not bind or cross link
CC	GM-1. They may be used in medicine as an immunomodulator or adjuvant.
CC	They may also be used as an inhibitor for toxin-induced diarrhoea.
CC	Therefore, the peptides may be used in the production of a composition
CC	for treating, preventing and/or modulating a disease associated with an
CC	immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463
CC	represent preferred peptides of the invention, AAY87460 being
CC	particularly preferred.
XX	
SQ	Sequence 21 AA;
	Query Match 94.4%; Score 101; DB 21; Length 21;
	Best Local Similarity 95.2%; Pred. No. 1.5e-10;
	Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY	1 GETFQVEVPGSQHDSQKAI 21
Db	1 gatfqvevpgsqhdsqkai 21
RESULT 14	
ID	AAY87462
XX	AAW04857 standard; Protein; 103 AA.
XX	
AC	AAW04857;
XX	

KW Toxin; subunit; vaccine; transgenic plant; immunogen; antigen;  
 KW adjuvant; immunisation.

OS Escherichia coli.

XX WO9612801-A1.

PN 02-MAY-1996.

PD 24-OCT-1995; 95WO-US13376.

PF 24-OCT-1994; 94US-0328716.

PR (TULA ) TULANE EDUCATIONAL FUND.

PA (TEXA ) UNIV TEXAS A & M SYSTEM.

XX Arntzen CJ, Clements JD, Haq TA, Mason HS;

PI WPI; 1996-230602/23.

DR N-PSDB; AAT18799, AAT18800.

XX Transgenic plants contg. E. coli heat labile enterotoxin subunits  
 PT used as oral vaccines for animals which consume the plant

PS Disclosure; Page 100-101; 130pp; English.

CC A transgenic plant comprising or expressing a DNA sequence encoding  
 CC an immunogenic agent can be used as an oral vaccine for animals.  
 CC The vaccine is administered by the oral consumption of the plant and  
 CC provides the first known functional method for immunising animals  
 CC using transgenic plants, where the plants express bacterial antigens  
 CC that act as both immunogens and adjuvants. The method provides an  
 CC inexpensive production and delivery system for such antigens to  
 CC animals. This is the LT-B Escherichia coli toxin subunit and its  
 CC coding sequence was used in the construction of such a transgenic  
 CC plant. The immunogenic agent preferably comprises the LT-B or CT-B  
 CC (cholera toxin B subunit) or optionally LT-A or CT-A.

XX Sequence 103 AA;

Query Match 94.4%; Score 101; DB 17; Length 103;  
 Best Local Similarity 95.2%; Pred. No. 9.7e-10;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSQHIDSQKKAI 21

Db 45 gatfgvevpgsqhidsqkai 65

RESULT 16

AAW06606

ID AAW06606 standard; Protein; 103 AA.

XX AAW06606;

AC 06-AUG-1997 (first entry)

XX Cholera toxin B subunit/heat labile enterotoxin B subunit hybrid.

XX Heat labile; tolerant; stable; hybrid; cholera toxin B subunit; CTB;  
 KW heat labile enterotoxin B subunit; LT-B; vaccine; immunisation;  
 KW immunogen; antigen; broad spectrum; cross reactive; diarrhoea;  
 KW vomiting; food poisoning.

XX Chimeric - Vibrio cholerae.

OS Chimeric - Enterotoxigenic Escherichia Coli.

XX Key Location/Qualifiers

FT Misc-difference 1

FT /label= substitution

FT /note= "wild-type Thr replaced by Ala"

FT Misc-difference 94

FT /label= substitution  
 FT /note= "wild-type His replaced by Asn"

FT Misc-difference 95

FT /label= substitution

FT /note= "wild-type Ala replaced by Ser"

XX WO9634893-A1.

PN 07-NOV-1996.

PD 02-MAY-1996; 96WO-SE00570.

PF 05-MAY-1995; 95SE-0001682.

PR (HOLM/) HOLMGREN J.

PA (LEBE/) LEBENS M R.

XX Holmgren J, Lebens MR;

PI WPI; 1996-506108/50.

DR N-PSDB; AAT43576.

XX Heat-labile enterotoxin B sub-unit and cholera toxin B sub-unit  
 PT hybrid protein - opt. fused to immunogenic sequence for use in  
 PT vaccines against enterotoxin-induced illness

PS Claim 3; Page -; 32pp; English.

XX AAW06606 is a mature cholera toxin B subunit (CTB)/heat labile  
 CC enterotoxin B subunit (LTB) hybrid molecule. In mature CTB protein,  
 CC certain amino acids (aa) were replaced with corresponding aa from  
 CC heat-labile enterotoxin B subunit (LTB). The specific amino acid  
 CC substitutions impart LTB-specific epitope characteristics to  
 CC immunogenic mature CTB. The hybrid molecules have increased  
 CC cross-reactivity and are suitable for a broad spectrum vaccine to  
 CC protect against enterotoxigenic illness. Immunogenic proteins  
 CC comprising the hybrid molecules can be used to treat, or in a  
 CC vaccine to prevent, enterotoxin-induced illness, e.g. diarrhoea  
 CC and vomiting, in humans and animals.

CC Note - This sequence does not appear in the specification, it is  
 CC a claimed mutant sequence of mature cholera toxin B subunit (see  
 CC AAW06605).

XX Sequence 103 AA;

Query Match 94.4%; Score 101; DB 17; Length 103;  
 Best Local Similarity 95.2%; Pred. No. 9.7e-10;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSQHIDSQKKAI 21

Db 45 gatfgvevpgsqhidsqkai 65

RESULT 17

AAW06607

ID AAW06607 standard; Protein; 103 AA.

XX AAW06607;

AC 06-AUG-1997 (first entry)

XX Cholera toxin B subunit/heat labile enterotoxin B subunit hybrid.

XX Heat labile; tolerant; stable; hybrid; cholera toxin B subunit; CTB;  
 KW heat labile enterotoxin B subunit; LTB; vaccine; immunisation;  
 KW immunogen; antigen; broad spectrum; cross reactive; diarrhoea;  
 KW vomiting; food poisoning.

XX Chimeric - Vibrio cholerae.

OS Chimeric - Enterotoxigenic Escherichia Coli.

XX



```

FH Key          Location/Qualifiers
FT Misc-difference 1..25
FT /label= substitution
FT /note= "the first 25 amino acids of mature
FT wild-type cholera toxin B subunit are
FT replaced with the first 25 amino acids
FT of mature enterotoxin B subunit"
XX
XX WO9634893-A1.
XX
XX 07-NOV-1996.
XX
XX 02-MAY-1996; 96WO-SE00570.
XX
XX 05-MAY-1995; 95SE-0001682.
XX
XX (HOLM/) HOLMGREN J.
XX (LEBE/) LEBENS M R.
XX
XX Holmgren J, Lebens MR;
XX
XX WPI; 1996-506108/50.
XX N-PSDB; AAT43577.
XX
XX Heat-labile enterotoxin B sub-unit and cholera toxin B sub-unit
XX hybrid protein - opt. fused to immunogenic sequence for use in
XX vaccines against enterotoxin-induced illness
XX
XX Claim 4; Page -; 32pp; English.
XX
XX AA06607 is a mature cholera toxin B subunit (CTB)/heat labile
XX enterotoxin B subunit (LTB) hybrid molecule. In mature CTB protein,
XX certain amino acids (aa) were replaced with corresponding aa from
XX heat-labile enterotoxin B subunit (LTB). The specific amino acid
XX substitutions impart LTB-specific epitope characteristics to
XX immunogenic mature CTB. The hybrid molecules have increased
XX cross-reactivity and are suitable for a broad spectrum vaccine to
XX protect against enterotoxigenic illness. Immunogenic proteins
XX comprising the hybrid molecules can be used to treat, or in a
XX vaccine to prevent, enterotoxin-induced illness, e.g. diarrhoea
XX and vomiting, in humans and animals.
XX Note - This sequence does not appear in the specification, it is
XX a claimed mutant sequence of mature cholera toxin B subunit (see
XX AA06605).
XX
XX Sequence 103 AA;

Query Match          94.4%; Score 101; DB 17; Length 103;
Best Local Similarity 95.2%; Pred. No. 9.7e-10;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETFQVEVPGSQHDSQKKAI 21
Db 45 gatfvevpgsqhidsqkkai 65

RESULT 18
AAW80808
ID AAW80808 standard; protein; 103 AA.
XX
XX AAW80808;
XX
XX 29-JAN-1999 (first entry)
XX
XX Amino acid sequence of the wild type chorela toxin B subunit.
XX
XX Cholera toxin B subunit; nontoxic subunit; adjuvant; coadministration;
XX antigen; bird; animal; mucosal; vaccine.
XX
XX Vibrio cholerae.
XX
XX WO9845324-A1.
XX

```

```

XX 15-OCT-1998.
XX
XX 03-APR-1998; 98WO-US06725.
XX
XX 04-APR-1997; 97US-0043410.
XX
XX (KIYO/) KIYONO H.
XX (MCGH/) MCGHEE J R.
XX (TAKE/) TAKEDA Y.
XX (UABR-) UAB RES FOUND.
XX (YAMA/) YAMAMOTO S.
XX
XX Kiyono H, Mcghee JR, Takeda Y, Yamamoto S;
XX WPI; 1998-594478/50.
XX
XX New mutant cholera toxin selected from a group comprising nontoxic
XX subunits/derivatives - effective as an adjuvant when coadministered
XX with an antigen to birds and mammals
XX
XX Disclosure; Fig 1B; 43pp; English.
XX
XX This is the amino acid sequence of the cholera toxin B subunit used in
XX the method of the invention involving the use of nontoxic subunits as
XX an effective adjuvant in coadministration of an antigen to birds
XX and animals. In addition to the use of the toxin as an mucosal
XX adjuvant, it also provides a vaccine comprising the toxin, an
XX immunogenic amount of an antigen, and a pharmaceutically acceptable
XX carrier. The toxin can be used with single/multiple vaccines, and it
XX enables the possibility for commercial mucosal adjuvants for use in
XX humans, since these are more effective and safer than vaccines
XX administered subcutaneously.
XX
XX Sequence 103 AA;

Query Match          94.4%; Score 101; DB 19; Length 103;
Best Local Similarity 95.2%; Pred. No. 9.7e-10;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETFQVEVPGSQHDSQKKAI 21
Db 45 gatfvevpgsqhidsqkkai 65

RESULT 19
AAR04163
ID AAR04163 standard; protein; 118 AA.
XX
XX AAR04163;
XX
XX 10-SEP-1990 (first entry)
XX
XX Cholera Toxin B-subunit.
XX
XX cholera toxin B-subunit; hybrid protein; heterologous IgA active antigen.
XX synthetic.
XX
XX Key          Location/Qualifiers
XX misc-difference 18..18
XX /*label= His or Tyr
XX region 1..11
XX /*label= signal peptide
XX /*note= absent from mature protein
XX
XX WO9003437-A.
XX
XX 05-APR-1990.
XX
XX 27-SEP-1989; 89WO-0000495.
XX

```

PR 27-SEP-1988; 88FR-0012627.  
 PA (UULI-) L'UNIVERSITE DE L'ETAT A LIEGE.  
 XX L'Hoir C, Renard A, Martial J;  
 PI WPI; 1990-132273/17.  
 DR N-NSDB; Q04046.  
 XX New hybrid protein, useful in vaccines -  
 PT contains cholera toxin B subunit and heterologous IgA active  
 PT antigenic sequence.  
 XX  
 PS Disclosure; ; pp; French.  
 CC Mature cholera toxin B-subunit is obtained when the signal peptide is  
 CC cleaved off. There is an ochre codon at position 343-5; the sequence  
 CC downstream from it is part of a plasmid.  
 XX  
 SQ Sequence 118 AA;  
 Query Match 94.4%; Score 101; DB 11; Length 118;  
 Best Local Similarity 95.2%; Pred. No. 1.1e-09;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GETFOVEVPGSQHDSOKKAI 21  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 56 gatfqvevpgsqhdsqkka 76  
 RESULT 20  
 AAP93561  
 ID AAP93561 standard; protein; 124 AA.  
 XX  
 AC AAP93561;  
 XX  
 DT 06-JUN-1990 (first entry)  
 XX  
 DE B subunit of the heat-labile enterotoxin (LT-B) derived from E. coli.  
 XX  
 KW B subunit; heat-labile enterotoxin; LT-B; Escherichia coli; malaria;  
 KW circumsporozoite protein; fusion protein; live recombinant vaccine;  
 KW Salmonella; epitope.  
 XX  
 OS Escherichia coli.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..22  
 FT /note="Signal peptide"  
 FT Protein 23..124  
 FT /note="Mature LT-B"  
 XX  
 PN W08902924-A.  
 XX  
 PD 06-APR-1989.  
 XX  
 PF 30-SEP-1988; 88WO-US03376.  
 XX  
 PR 02-OCT-1987; 87US-0104735.  
 XX  
 PA (PRAX-) PRAXIS BIOLOGICS IN.  
 XX  
 PI Brey RN, Majarian WR, Pillai S, Hockmeyer WT;  
 XX  
 DR WPI; 1989-114399/15.  
 DR N-PSDB; AAN90747.  
 XX  
 XX Live recombinant vaccine for malaria -  
 PT comprising attenuated entero-invasive bacterium contg. DNA  
 PT encoding epitope of malaria parasite  
 XX  
 PS Fig 3; p. 3/17; 105pp; English.

XX  
 CC In the patent, the DNA encoding LT-B is expressed as part of a fusion  
 CC protein with an epitope of a malaria parasite, eg Region I or Region II  
 CC or a repeat region of circumsporozoite protein antigen (CS) (AAP93560)  
 CC from plasmodium berghel. Pref. the fusion gene is inserted into  
 CC attenuated Salmonella enteritidis under the left promoter control of  
 CC lambda. Such bacteria can multiply in the host without causing disease or  
 CC disorder and express CS that will induce a protective immune response  
 CC against malaria and can be used in vaccines. Such vaccines can be  
 CC multivalent.  
 XX  
 SQ Sequence 124 AA;  
 Query Match 94.4%; Score 101; DB 10; Length 124;  
 Best Local Similarity 95.2%; Pred. No. 1.2e-09;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GETFOVEVPGSQHDSOKKAI 21  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 56 gatfqvevpgsqhdsqkka 86  
 RESULT 21  
 AAW06605  
 ID AAW06605 standard; Protein; 124 AA.  
 XX  
 AC AAW06605;  
 XX  
 DT 06-AUG-1997 (first entry)  
 XX  
 DE Cholera toxin B subunit, used for hybrid immunogenic toxin production.  
 XX  
 KW Heat labile; tolerant; stable; hybrid; cholera toxin B subunit; CTB;  
 KW heat labile enterotoxin B subunit; LTb; vaccine; immunisation;  
 KW immunogen; antigen; broad spectrum; cross reactive; diarrhoea;  
 KW vomiting; food poisoning.  
 XX  
 OS Vibrio cholerae.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /label= sig\_peptide  
 FT Protein 22..124  
 FT /label= mat\_protein  
 XX  
 PN W09634893-A1.  
 XX  
 PD 07-NOV-1996.  
 XX  
 PF 02-MAY-1996; 96WO-SE00570.  
 XX  
 PR 05-MAY-1995; 95SE-0001682.  
 XX  
 PA (HOLM/) HOLMGREN J.  
 PA (LEBE/) LEBENS M R.  
 XX  
 PI Holmgren J, Lebens MR;  
 XX  
 DR WPI; 1996-506108/50.  
 DR N-PSDB; AAT43575.  
 XX  
 PT Heat-labile enterotoxin B sub-unit and cholera toxin B sub-unit  
 PT hybrid protein - opt. fused to immunogenic sequence for use in  
 PT vaccines against enterotoxin-induced illness  
 XX  
 PS Disclosure; Fig 1; 32pp; English.  
 XX  
 CC AAW06605 is the full (including the signal peptide) length cholera  
 CC toxin B subunit (CTB), this sequence is described as unpublished in  
 CC the specification. The mature CTB protein was used to create hybrid  
 CC mutants, in which certain amino acids (aa) of CTB were replaced with  
 CC corresponding aa from heat-labile enterotoxin B subunit (LTB), see

Query Match	94.48:	Score 101:	DB 19:	Length 124:
-------------	--------	------------	--------	-------------

CC This allows the use of a strategically placed EcoRI site just upstream  
CC of the RBS on the LTB gene for the insertion of a strong tac promoter  
CC for the expression of CMB. The protein can be used as vaccines,  
CC diagnostic reagents and receptor-blocking agents for prophylaxis of  
CC cholera and E.coli diarrhoea.  
XX  
XX  
SO Sequence 131 AA:

Query Match 94.4%; Score 101; DB 11; Length 131;  
Best Local Similarity 95.2%; Pred. No. 1.3e-09;  
Matches 20: Conservative 0; Mismatches 1; Indels 0; Gaps

**Qy**      1 GETFQVEVPGSHIDSQKAI 21  
         | |||||  
**Db**      73 gatfovevoagshidsqkai 93

REC- 22  
AAR50227

XX  
-----

XX

XX  
XXXX  
1  
1

KW Group A streptococci; rheumatic fever; pharyngitis.

os synthetic.

PN WO9406465-A.

PD 31-MAR-1994.

PF . 15-SEP-1993; 93WO-US08704.

PR 16-SEP-1992; 92US-0945860.

PA (UYTE-) UNIV TENNESSEE RES CORP.

PI Dale JB;

DR WPI; 1994-118162/14.

XX

PT elicit opsonic antibodies without eliciting cross-reactive

XX XX

XX

CC factor and protective antigen of these organisms. However, there are

CC recombinant M protein antigens comprising a gene encoding a carrier

CC or more epitopes. The carrier may be the B subunit of E.coli labile

CC eg AAR50226. The LT-B-M24 fusion gene of the example was expressed

CC oligos which copied the first 36 bp of the emm 24 gene. Rabbits

CC antibodies against type 24 streptococci.

SQ	Sequence	138 AA;
SQ	Sequence	138 AA;

Best Local Similarity 95.2%; Pred. No. 1.4e-09;

Search completed: July 16, 2001, 16:35:42  
Job time: 206 sec

***This Page Blank (uspto)***

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 16:32:16 ; Search time 56.06 Seconds  
(Without alignments)  
7.570 Million cell updates/sec

Title: US-09-786-648-2  
Perfect score: 39  
Sequence: 1 EYPGSQH 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_0601.\*

```
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	100.0	7	21 AAY87460	Cholera toxin B/en
2	39	100.0	8	21 AAY87461	Cholera toxin B/en
3	39	100.0	12	21 AAY87461	Cholera toxin B/en
4	39	100.0	15	10 AAP93498	CTP3 epitope of th
5	39	100.0	15	16 AAR85125	Cholera toxin B an
6	39	100.0	21	21 AAY87462	Cholera toxin B/en
7	39	100.0	21	21 AAY87463	E. coli heat labil
8	39	100.0	23	16 AAR76748	Residues 50-64 of
9	39	100.0	26	4 AAP30265	Sequence of amino
10	39	100.0	41	6 AAP50439	Network polymer wh
11	39	100.0	46	6 AAP50436	Network polymer wh

12	39	100.0	47	4 AAP30600	Sequence of amino
13	39	100.0	93	16 AAR72545	ADP-ribosylating t
14	39	100.0	93	20 AAY41816	Escherichia coli v
15	39	100.0	93	20 AAW95226	E. coli heat-labil
16	39	100.0	93	21 AAY68365	Heat labile toxin
17	39	100.0	93	22 AAB66239	E coli verotoxin-1
18	39	100.0	103	6 AAP50340	Sequence of sub-un
19	39	100.0	103	17 AAW04857	Synthetic cholera
20	39	100.0	103	17 AAR94939	Heat labile entero
21	39	100.0	103	17 AAW06606	Cholera toxin B su
22	39	100.0	103	17 AAW06607	Cholera toxin B su
23	39	100.0	103	19 AAW80808	Amino acid sequenc
24	39	100.0	118	11 AAR04163	Cholera Toxin B-su
25	39	100.0	124	11 AAR93561	B subunit of the h
26	39	100.0	124	17 AAW06605	Cholera toxin B su
27	39	100.0	124	19 AAW59770	Amino acid sequenc
28	39	100.0	124	21 AAY96652	Plant-optimized E.
29	39	100.0	124	21 AAY96872	Plant-optimized V.
30	39	100.0	126	12 AAR12630	GtFB.l/CTB chimeri
31	39	100.0	131	11 AAR04825	LTB-CTB fusion pro
32	39	100.0	134	22 AAB73241	Recombinant exotox
33	39	100.0	138	15 AAR50227	Sequence of LT-B-M
34	39	100.0	142	22 AAB73242	Recombinant exotox
35	39	100.0	155	22 AAB73243	Recombinant exotox
36	39	100.0	163	22 AAB73244	Recombinant exotox
37	39	100.0	170	20 AAW94082	LTB-CTP fusion pro
38	39	100.0	371	20 AAY01300	Labile toxin (LT-B
39	39	100.0	371	20 AAW67443	C. jejuni flagelli
40	39	100.0	405	12 AAR11272	HSV-1 antigen/heat
41	34	87.2	51	21 AAG02635	Human secreted pro
42	33	84.6	124	13 AAR28831	B subunit of CT
43	33	84.6	318	21 AAY91645	Human secreted pro
44	33	84.6	379	19 AAW80937	Human heart lecith
45	33	84.6	379	19 AAW80939	Murine lecithin-ch

#### ALIGNMENTS

#### RESULT 1

AA1987460  
ID AAY87460 standard; peptide; 7 AA.

AC AAY87460;

XX 03-JUL-2000 (first entry)

DT Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:2.

DE Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;  
KW beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;  
KW adjuvant; immune disorder; diarrhoea.

XX Vibrio cholerae.

OS Escherichia coli.

XX WO200014114-A1.

PD 16-MAR-2000.

XX 07-SEP-1999; 99WO-GB02970.

XX 07-SEP-1998; 98GB-0019484.

XX (UYBR-) UNIV BRISTOL.

XX Williams NA, Hirst TR;

XX WPI; 2000-256943/22.

XX Derivatives of Escherichia coli heat labile enterotoxins useful as  
immunomodulators and for treating diarrhea and which do not bind the  
glycolipid receptor GM-1 -

XX PS Claim 1; Page 13; 62pp; English.

XX CC The invention relates to peptide fragments of the *Escherichia coli* heat

XX CC labile enterotoxin (Etx) and its closely related homologue, cholera

XX CC toxin (Ctx) from *Vibrio cholerae* which do not bind to the ubiquitous

XX CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are

XX CC composed of one A subunit and five identical B subunits. The A subunit

XX CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-

XX CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)

XX CC facilitate the entry of subunit A into the host cell via the binding and

XX CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible

XX CC for some of the effects of Etx and Ctx, it has been found that certain

XX CC effects of the toxins, such as immunomodulation, are not mediated

XX CC through GM-1 binding. The peptides of the invention (AAV87460-Y87463) are

XX CC fragments of the beta-4-alpha-2 loop of EtxB and/or CtxB, exerting the

XX CC same effects as normal EtxB and CtxB subunits, except that they do not

XX CC bind or cross link GM-1. They may be used in medicine as an

XX CC immunomodulator or adjuvant. They may also be used as an inhibitor for

XX CC toxin-induced diarrhoea. Therefore, the peptides may be used in the

XX CC production of a composition for treating, preventing and/or modulating a

XX CC disease associated with an immune disorder and/or toxin-induced

XX CC diarrhoea. Sequences AAV87460-Y87463 represent peptides used in an

XX CC exemplification of the present invention to assess whether a peptide

XX CC corresponding to a portion of the beta-4-alpha-2 loop of EtxB and CtxB

XX CC has immunomodulatory effects. Peptide AAV87460 corresponds to residues

XX CC 51-58 of the EtxB/CtxB beta-4-alpha-2 loop, and peptide AAV87463 is a

XX CC randomly selected control peptide.

XX SQ Sequence 7 AA;

Query Match 100.0%; Score 39; DB 21; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSGH 7  
 |||||  
 Db 1 evpgsqh 7

RESULT 2

AAV87464

ID AAV87464 standard; peptide; 8 AA.

XX AC AAV87464;

XX DT 03-JUL-2000 (first entry)

XX DE Cholera toxin B/enterotoxin B-derived immunomodulatory peptide.

XX KW Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;  
 beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;  
 adjuvant; immune disorder; diarrhoea.

XX OS *Vibrio cholerae*.

XX OS *Escherichia coli*.

XX PN WO200014114-A1.

XX PD 16-MAR-2000.

XX PF 07-SEP-1999; 99WO-GB02970.

XX PR 07-SEP-1998; 98GB-0019484.

XX XX (UYBR-) UNIV BRISTOL.

XX PA Williams NA, Hirst TR;

XX PI WPI; 2000-256943/22.

XX XX Derivatives of *Escherichia coli* heat labile enterotoxins useful as  
 immunomodulators and for treating diarrhoea and which do not bind the  
 glycolipid receptor GM-1.

XX PS Example 5; Page 45; 62pp; English.

XX CC The invention relates to peptide fragments of the *Escherichia coli* heat

XX CC labile enterotoxin (Etx) and its closely related homologue, cholera

XX CC toxin (Ctx) from *Vibrio cholerae* which do not bind to the ubiquitous

XX CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are

XX CC composed of one A subunit and five identical B subunits. The A subunit

XX CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-

XX CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)

XX CC facilitate the entry of subunit A into the host cell via the binding and

XX CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible

XX CC for some of the effects of Etx and Ctx, it has been found that certain

XX CC effects of the toxins, such as immunomodulation, are not mediated

XX CC through GM-1 binding. The peptides of the invention (AAV87461-Y87463) are

XX CC fragments of the beta-4-alpha-2 loop of EtxB and/or CtxB, exerting the

XX CC same effects as normal EtxB and CtxB subunits, except that they do not

XX CC bind or cross link GM-1. They may be used in medicine as an

XX CC immunomodulator or adjuvant. They may also be used as an inhibitor for

XX CC toxin-induced diarrhoea. Therefore, the peptides may be used in the

XX CC production of a composition for treating, preventing and/or modulating a

XX CC disease associated with an immune disorder and/or toxin-induced

XX CC diarrhoea. Sequences AAV87464-Y87465 represent peptides used in an

XX CC exemplification of the present invention to assess whether a peptide

XX CC corresponding to a portion of the beta-4-alpha-2 loop of EtxB and CtxB

XX CC has immunomodulatory effects. Peptide AAV87464 corresponds to residues

XX CC 51-58 of the EtxB/CtxB beta-4-alpha-2 loop, and peptide AAV87465 is a

XX CC randomly selected control peptide.

XX SQ Sequence 8 AA;

Query Match 100.0%; Score 39; DB 21; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSGH 7  
 |||||  
 Db 1 evpgsqh 7

RESULT 3

AAV87461

ID AAV87461 standard; peptide; 12 AA.

XX AC AAV87461;

XX DT 03-JUL-2000 (first entry)

XX DE Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:3.

XX KW Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;  
 beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;  
 adjuvant; immune disorder; diarrhoea.

XX OS *Vibrio cholerae*.

XX OS *Escherichia coli*.

XX PN WO200014114-A1.

XX PD 16-MAR-2000.

XX PF 07-SEP-1999; 99WO-GB02970.

XX PR 07-SEP-1998; 98GB-0019484.

XX XX (UYBR-) UNIV BRISTOL.

XX PA Williams NA, Hirst TR;

XX PI WPI; 2000-256943/22.

XX XX Derivatives of *Escherichia coli* heat labile enterotoxins useful as  
 immunomodulators and for treating diarrhoea and which do not bind the  
 glycolipid receptor GM-1.



XX Disclosure; Page 15; 62pp; English.  
 XX The invention relates to peptide fragments of the *Escherichia coli* heat  
 CC labile enterotoxin (Etx) and its closely related homologue, cholera  
 CC toxin (Ctx) from *Vibrio cholerae* which do not bind to the ubiquitous  
 CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are  
 CC composed of one A subunit and five identical B subunits. The A subunit  
 CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-  
 CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)  
 CC facilitate the entry of subunit A into the host cell via the binding and  
 CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible  
 CC for some of the effects of Etx and Ctx, it has been found that certain  
 CC effects of the toxins, such as immunomodulation, are not mediated  
 CC through GM-1 binding. The peptides of the invention are fragments of the  
 CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as  
 CC normal EtxB and CtxB subunits, except that they do not bind or cross link  
 CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.  
 CC They may also be used as an inhibitor for toxin-induced diarrhoea.  
 CC Therefore, the peptides may be used in the production of a composition  
 CC for treating, preventing and/or modulating a disease associated with an  
 CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463  
 CC represent preferred peptides of the invention, AAY87460 being  
 CC particularly preferred.  
 XX Sequence 12 AA;

Query Match 100.0%; Score 39; DB 21; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 0.16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
 DB 2 evpgsqh 8  
 |||||

RESULT 4  
 ARP93498  
 ID AAP93498 standard; protein; 15 AA.

XX AAP93498;  
 AC  
 DT 03-MAY-1990 (first entry)  
 XX CTP3 epitope of the Cholera toxin B subunit.

DE CTP3 epitope of the Cholera toxin B subunit; flagellin fusion protein;  
 KW vaccine; immunotherapy; ds;  
 XX WO8910967-A.

PN 16-NOV-1989.

PD 05-MAY-1989; 89WO-US01932.

PF 05-MAY-1988; 88US-0190570.

XX (PRAX-) PRAXIS BIOLOGICS INC.

PA (STRO) LEYLAND STANFORD JUNIOR UNIV.  
 XX Marjarian WR, Stocker BAD, Newton SMC;

XX WPI; 1989-356496/48.

DR N-PSDB; AAN92414.  
 XX New recombinant flagellin gene including sequence - for heterologous  
 PT epitope, and expressed fusion proteins, useful in vaccines and for prodn.  
 PT of antibodies.

XX Disclosure; fig.4B; 137pp; English.

PS This sequence corresponds to the CTP3 epitope of the Cholera toxin B

CC

CC subunit. The DNA sequence encoding this ligates to othersynthetic  
 CC oligonucleotides to form a new recombinant gene. This encodes  
 CC a flagellin fusion protein which can be used in vaccines for immuno-  
 CC therapy.

SQ Sequence 15 AA;

Query Match 100.0%; Score 39; DB 10; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 0.2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
 DB 2 evpgsqh 8  
 |||||

RESULT 5  
 AAR85125  
 ID AAR85125 standard; peptide; 15 AA.

XX AAR85125;

AC 13-JUN-1996 (first entry)

DE Cholera toxin B antigenic peptide fragment CTP3.

XX Conjugate; cholera; B toxin; peptide fragment; microparticulate;  
 KW inert carrier; modified silica; thyroglobulin; oral vaccine;  
 KW immunisation; infection; insoluble; digestive tract; antigen;  
 KW intestines; antibodies; secretory; IgA class.

XX *Vibrio cholerae*.

OS WO9529701-A1.

PN 09-NOV-1995.

XX 02-MAY-1995; 95WO-EP01661.

XX 03-MAY-1994; 94IL-0109519.

XX (YEDA) YEDA RES & DEV CO LTD.

XX Marks RS, Mirelman D, Sela M;

PI WPI; 1995-403805/51.

XX Vaccines for oral immunisation against infecting agents, e.g.  
 PT cholera - comprise a conjugate of an antigen of an infecting agent  
 PT covalently bound to micro:particulate inert carrier, e.g. modified  
 PT aldehyde silica

PS Claim 7; Page 25; 40pp; English.

XX A compsn. comprising a conjugate of an antigenic cholera B toxin  
 CC peptide fragment (e.g. AAR85126-30 and esp. AAR85125), covalently  
 CC bound to a microparticulate inert carrier (e.g. modified silica or  
 CC thyroglobulin) can be used as an oral vaccine for immunisation  
 CC against cholera infection. The inert carrier is insoluble in the  
 CC digestive tract, allowing presentation of the antigen in the  
 CC intestines, where it will elicit antibodies mainly of the  
 CC secretory IgA class.

SQ Sequence 15 AA;

Query Match 100.0%; Score 39; DB 16; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 0.2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
 |||||

Db 2 evpgsqh 8

## RESULT 6

AA87462  
ID AAY87462 standard; peptide; 21 AA.

XX AC AAY87462;  
XX DT 03-JUL-2000 (first entry)

XX DE Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:4.

XX KW Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;

XX KW beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;  
XX KW adjuvant; immune disorder; diarrhoea.

XX OS Vibrio cholerae.  
XX OS Escherichia coli.

XX PN WO200014114-A1.

XX PD 16-MAR-2000.

XX PF 07-SEP-1999; 99WO-GB02970.

XX PR 07-SEP-1998; 98GB-0019484.

XX PA (UYBR-) UNIV BRISTOL.

XX PI Williams NA, Hirst TR;

XX DR WPI; 2000-256943/22.

XX PT Derivatives of Escherichia coli heat labile enterotoxins useful as  
XX PT immunomodulators and for treating diarrhea and which do not bind the  
XX PT glycolipid receptor GM-1 -

XX PS Disclosure; Page 15; 62pp; English.

XX CC The invention relates to peptide fragments of the Escherichia coli heat  
XX CC labile enterotoxin (Etx) and its closely related homologue, cholera  
XX CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous  
XX CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are  
XX CC composed of one A subunit and five identical B subunits. The A subunit  
XX CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-  
XX CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)  
XX CC facilitate the entry of subunit A into the host cell via the binding and  
XX CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible  
XX CC for some of the effects of Etx and Ctx, it has been found that certain  
XX CC effects of the toxins, such as immunomodulation, are not mediated  
XX CC through GM-1 binding. The peptides of the invention are fragments of the  
XX CC beta-4-alpha-2 loop of EtxB and/or CtxB, exerting the same effects as  
XX CC normal EtxB and CtxB subunits, except that they do not bind or cross link  
XX CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.  
XX CC They may also be used as an inhibitor for toxin-induced diarrhoea.  
XX CC Therefore, the peptides may be used in the production of a composition  
XX CC for treating, preventing and/or modulating a disease associated with an  
XX CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463  
XX CC represent preferred peptides of the invention, AAY87460 being  
XX CC particularly preferred.

XX SQ Sequence 21 AA;

Query Match 100.0%; Score 39; DB 21; Length 21;  
Best Local Similarity 100.0%; Pred. NO. 0.28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVPSQH 7

Db 7 evpgsqh 13

## RESULT 8

AA87463

## RESULT 7

AA87463

XX ID AAY87463 standard; peptide; 21 AA.

XX AC AAY87463;

XX DT 03-JUL-2000 (first entry)

XX DE E. coli heat labile enterotoxin B-derived peptide, SEQ ID NO:5.

XX KW Heat labile enterotoxin subunit B; EtxB;

XX KW beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;  
XX KW adjuvant; immune disorder; diarrhoea.

XX OS Escherichia coli.

XX PN WO200014114-A1.

XX PD 16-MAR-2000.

XX PF 07-SEP-1999; 99WO-GB02970.

XX PR 07-SEP-1998; 98GB-0019484.

XX PA (UYBR-) UNIV BRISTOL.

XX PI Williams NA, Hirst TR;

XX DR WPI; 2000-256943/22.

XX PT Derivatives of Escherichia coli heat labile enterotoxins useful as  
XX PT immunomodulators and for treating diarrhea and which do not bind the  
XX PT glycolipid receptor GM-1 -

XX PS Disclosure; Page 15; 62pp; English.

XX CC The invention relates to peptide fragments of the Escherichia coli heat  
XX CC labile enterotoxin (Etx) and its closely related homologue, cholera  
XX CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous  
XX CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are  
XX CC composed of one A subunit and five identical B subunits. The A subunit  
XX CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-  
XX CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)  
XX CC facilitate the entry of subunit A into the host cell via the binding and  
XX CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible  
XX CC for some of the effects of Etx and Ctx, it has been found that certain  
XX CC effects of the toxins, such as immunomodulation, are not mediated  
XX CC through GM-1 binding. The peptides of the invention are fragments of the  
XX CC beta-4-alpha-2 loop of EtxB and/or CtxB, exerting the same effects as  
XX CC normal EtxB and CtxB subunits, except that they do not bind or cross link  
XX CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.  
XX CC They may also be used as an inhibitor for toxin-induced diarrhoea.  
XX CC Therefore, the peptides may be used in the production of a composition  
XX CC for treating, preventing and/or modulating a disease associated with an  
XX CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463  
XX CC represent preferred peptides of the invention, AAY87460 being  
XX CC particularly preferred.

XX SQ Sequence 21 AA;

Query Match 100.0%; Score 39; DB 21; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVPSQH 7

Db 7 evpgsqh 13

ID AAR76748 standard; Protein; 23 AA.  
 AC AAR76748;  
 XX  
 XX 18-MAR-1996 (first entry)  
 XX  
 XX Residues 50-64 of cholera toxin B subunit and FimH 224-226.  
 XX  
 XX FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;  
 KW FimA; FimF; FimG; receptor binding site; PCR; amplify; ss.  
 XX  
 XX Chimeric - Vibrio cholerae.  
 OS Chimeric - Escherichia coli.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1 /note= "Represents FimH residue 224"  
 FT Peptide 2..4 /note= "Linker peptide"  
 FT Peptide 5..19 /note= "Cholera toxin B subunit 50-64"  
 FT Peptide 20..22 /note= "Linker peptide"  
 FT Misc-difference 23 /note= "Represents FimH residue 226"  
 FT  
 XX W09520657-Al.  
 PN  
 XX  
 PD 03-AUG-1995.  
 XX  
 XX 27-JAN-1995; 95WO-DK00042.  
 PF  
 XX 27-JAN-1994; 94US-0187166.  
 PR  
 XX (GXBI-) GX BIOSYSTEMS AS.  
 PA  
 XX Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;  
 PI  
 XX WPI; 1995-275442/36.  
 DR N-PSDB; AAQ93061.  
 XX  
 XX Receptor specific bacterial adhesins - useful for targetting active  
 PT compounds and microbial cells to locations of receptors  
 XX  
 XX Example 3; Page 58; 152pp; English.  
 XX  
 XX This sequence is encoded by a fragment of the the plasmid pLPA93  
 CC which was used in the production of fimH fusion genes comprising  
 CC the cholera toxin B subunit inserted into the fimH gene. This insert  
 CC shows the inclusion of the B subunit into the FimH protein at position  
 CC 224-226. The chimeric genes were then opt. further modified by insertion  
 CC of the hepatitis B virus surface antigen pre-S2 region into a different  
 CC position of the FimH adhesin of type 1 fimbriae. Restriction site handles  
 CC (BglII-sites) were introduced into the fimH gene, and the foreign  
 CC epitopes are then inserted in-frame. In the selected positions the  
 CC insertion of the epitopes did not significantly alter the adhesive  
 CC function of the FimH protein. The expression of the chimeric proteins  
 CC on the surface of fimbriae on bacterial hosts illustrated the possibility  
 CC of using bacterial adhesins as general presenters of foreign antigens and  
 CC epitopes. These chimeric genes may be used in the production of variant  
 CC FimH adhesins which may be useful for targetting active compounds  
 CC and microbial cells to locations comprising selected receptors to which  
 CC the adhesins bind.  
 XX  
 XX Sequence 23 AA;  
 SQ  
 Query Match 100.0%; Score 39; DB 16; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 0.31;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EYPGSOH 7  
 |||||

Db 6 evpgsqh 12  
 RESULT 9  
 AAP30265  
 ID AAP30265 standard; Protein; 26 AA.  
 XX  
 XX AAP30265;  
 AC  
 XX 21-APR-1992 (first entry)  
 DT  
 XX Sequence of amino acids 50-75 of the cholera toxin B1 subunit which  
 DE carries an Arg at posns. 67 and 73.  
 DE  
 XX Cholera vaccine; therapy; E.coli infection; enterotoxin LT.  
 KW  
 XX Vibrio cholerae.  
 OS  
 XX EP95426-A.  
 PN  
 XX 30-NOV-1983.  
 PD  
 XX 26-MAY-1983; 83EP-0401052.  
 PF  
 XX 26-MAY-1982; 82FR-0009167.  
 PR  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 PA (INSP ) INST PASTEUR.  
 PA  
 XX Milhaud G, Raulais D, Rivaille P, Siffert O, Dodin A;  
 PI Guyongruaz A, Delmas A;  
 PI  
 XX WPI; 1983-834645/49.  
 DR  
 XX  
 XX Cholera toxin B, sub-unit polypeptide(s) as vaccines and  
 PT medicaments - effective against Escherichia coli and Vibrio  
 PT cholerae infections, are prepd. by solid phase peptide synthesis  
 PT  
 XX Claim 7; Page 11; 13pp; French.  
 PS  
 XX The inventors claim cholera toxin B1 subunit sequences which carry  
 CC Arg in posn. 35, 67 or 73 (AAP30264, AAP30265, AAP30600). Tests have  
 CC shown that Arg in posns. 35, 67 and 73 play an important role in fixing  
 CC these toxins to cell walls. The peptides are used in the treatment of,  
 CC and vaccination against cholera infections and animal and human  
 CC infections due to E. coli (enterotoxin LT). The medicament may be  
 CC administered by oral, intraperitoneal, sub-cutaneous or intravenous  
 CC routes. For vaccines, pref. peptides having 15-30 AAs are injected  
 CC without carriers. Unit dose when used as a medicament is 50-500mg as  
 CC a vaccine 1-10mg of active cpd.  
 XX  
 XX Sequence 26 AA;  
 SQ  
 Query Match 100.0%; Score 39; DB 4; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 0.35;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EYPGSOH 7  
 |||||  
 Db 2 evpgsqh 8  
 RESULT 10  
 AAP50439  
 ID AAP50439 standard; protein; 41 AA.  
 XX  
 XX AAP50439;  
 AC  
 XX 01-JAN-1980 (first entry)  
 DT  
 XX Network polymer which comprises a series of composite E. coli heat-  
 DE labile toxin (LT)/heat-stable toxin (ST) polypeptide repeating units.

XX Heat-labile enterotoxin; heat-stable toxin; vaccine; ss.  
 KW Synthetic.  
 OS  
 PN W08502611-A.  
 XX  
 PD 20-JUN-1985.  
 XX  
 PF 12-DEC-1984; 84WO-US02030.  
 XX  
 PR 12-DEC-1983; 83US-0559469.  
 XX  
 PA (SCRI-) SCRIPPS CLINIC RES.  
 XX  
 PI Houghten RA;  
 XX  
 DR WPI; 1985-159230/26.  
 XX  
 PT New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -  
 useful for vaccination of mammals against the enterotoxin(s)  
 XX  
 PS Claim 8; Page 100; 120pp; English.  
 XX  
 CC The repeating units are bonded together by intramolecular  
 interpolypeptide cystine bonds formed between oxidized Cys residues  
 of the repeating units. This polypeptide may be used in the  
 CC vaccination of mammals for protection against the enterotoxins. The  
 CC composite polypeptide is made by solid phase synthesis or  
 CC recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,  
 CC X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.  
 XX  
 SQ Sequence 41 AA;  
 Query Match 100.0%; Score 39; DB 6; Length 41;  
 Best Local Similarity 100.0%; Pred. No. 0.55;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EVPGSQH 7  
 DB 15 evpgsqh 21  
 RESULT 11  
 AAP50436  
 ID AAP50436 standard; protein; 46 AA.  
 XX  
 AC AAP50436;  
 XX  
 DT 01-JAN-1980 (first entry)  
 XX  
 DE Network polymer which comprises a series of composite E. coli heat-  
 DE labile toxin (LT)/heat-stable toxin (ST) polypeptide repeating units.  
 XX  
 KW Heat-labile enterotoxin; heat-stable toxin; vaccine; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN W08502611-A.  
 XX  
 PD 20-JUN-1985.  
 XX  
 PF 12-DEC-1984; 84WO-US02030.  
 XX  
 PR 12-DEC-1983; 83US-0559469.  
 XX  
 PA (SCRI-) SCRIPPS CLINIC RES.  
 XX  
 PI Houghten RA;  
 XX  
 DR WPI; 1985-159230/26.  
 XX

PT New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -  
 useful for vaccination of mammals against the enterotoxin(s)  
 XX  
 PS Claim 8; Page 100; 120pp; English.  
 XX  
 CC The repeating units are bonded together by intramolecular  
 interpolypeptide cystine bonds formed between oxidized Cys residues  
 of the repeating units. This polypeptide may be used in the  
 CC vaccination of mammals for protection against the enterotoxins. The  
 CC composite polypeptide is made by solid phase synthesis or  
 CC recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,  
 CC X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.  
 XX  
 SQ Sequence 46 AA;  
 Query Match 100.0%; Score 39; DB 6; Length 46;  
 Best Local Similarity 100.0%; Pred. No. 0.62;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EVPGSQH 7  
 DB 17 evpgsqh 23  
 RESULT 12  
 AAP30600  
 ID AAP30600 standard; Protein; 47 AA.  
 XX  
 AC AAP30600;  
 XX  
 DT 21-APR-1992 (first entry)  
 XX  
 DE Sequence of amino acids 350-75 of the cholera toxin B1 subunit which  
 DE carries an Arg at posns. 35, 67 and 73.  
 XX  
 KW Cholera vaccine; therapy; E.coli infection; enterotoxin LT.  
 XX  
 OS Vibrio cholerae.  
 XX  
 PN EP95426-A.  
 XX  
 PD 30-NOV-1983.  
 XX  
 PF 26-MAY-1983; 83EP-0401052.  
 XX  
 PR 26-MAY-1982; 82PR-0009167.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 PA (INSP ) INST PASTEUR.  
 XX  
 PI Milhaud G, Raulais D, Rivaille P, Siffert O, Dodin A;  
 PI Guyongruaz A, Delmas A;  
 XX  
 DR WPI; 1983-834645/49.  
 XX  
 PT Cholera toxin B, sub-unit polypeptide(s) as vaccines and  
 PT medicaments - effective against Escherichia coli and Vibrio  
 PT cholerae infections, are prepd. by solid phase peptide synthesis  
 XX  
 PS Claim 8; Page 11; 13pp; French.  
 XX  
 CC The inventors claim cholera toxin B1 subunit sequences which carry  
 CC Arg in posn. 35, 67 or 73 (AAP30264, AAP30265, AAP30600). Tests have  
 CC shown that Arg in posns. 35, 67 and 73 play an important role in fixing  
 CC these toxins to cell walls. The peptides are used in the treatment of,  
 CC and vaccination against cholera infections and animal and human  
 CC infections due to E. coli (enterotoxin LT). The medicament may be  
 CC administered by oral, intraperitoneal, sub-cutaneous or intravenous  
 CC routes. For vaccines, pref. peptides having 15-30 AAs are injected  
 CC without carriers. Unit dose when used as a medicament is 50-500mg as  
 CC a vaccine 1-10mg of active cpd.  
 XX

SQ Sequence 47 AA;

Query Match 100.0%; Score 39; DB 4; Length 47;  
 Best Local Similarity 100.0%; Pred. No. 0.64;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
 |||||  
 Db 23 evpgsqh 29

RESULT 13

AAR72545  
 ID AAR72545 standard; peptide; 93 AA.

XX AC  
 XX AAR72545;

XX DT 28-NOV-1995 (first entry)

XX DE ADP-ribosylating toxin (verotoxin-1 B-subunit).

XX KW ADP-ribosylating toxin; pertussis holotoxin; B-subunit;  
 KW active site; E. coli heat labile toxin; verotoxin-1;  
 KW Bordetella pertussis vaccines.

XX OS Bacteria sp.

XX PN EP646599-A.

XX PD 05-APR-1995.

XX PF 23-AUG-1994; 94EP-0306219.

XX PR 24-AUG-1993; 93US-0110947.

XX PR 31-MAY-1994; 94US-0251121.

XX PA (CONN-) CONNAUGHT LAB LTD.

XX PA (UYAL-) UNIV ALBERTA.

XX PI Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;  
 XX Oomen R, Read RJ, Stein PE;

XX DR WPI; 1995-132623/18.

XX PT New modified forms of pertussis holotoxin - developed using  
 PT crystalline forms of pertussis holotoxin and its complexes with  
 PT other molecules

XX PS Disclosure; Fig 5; 54pp; English.

XX CC AAR72540-R72545 are structurally equivalent B-subunits from three  
 CC ADP-ribosylating toxins, pertussis holotoxin (PT), E. coli heat  
 CC labile toxin (LT), and verotoxin-1 (VT). The structural  
 CC information obtd. from these comparisons was used to identify  
 CC sites which contribute to PT's biological activity. By modifying  
 CC these sites the claimed PT mutants of the invention were produced,  
 CC they can be used in the development of vaccines against Bordetella  
 CC pertussis infection.

SQ Sequence 93 AA;

Query Match 100.0%; Score 39; DB 16; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
 |||||  
 Db 41 evpgsqh 47

RESULT 14

AAW95226  
 ID AAW95226 standard; peptide; 93 AA.

AA41816

XX ID AAY41816 standard; peptide; 93 AA.

XX AC AAY41816;

XX DT 08-DEC-1999 (first entry)

XX DE Escherichia coli verotoxin-1 B-subunit.

XX KW ADP-ribosylating toxin; PT; crystalline pertussis holotoxin; ETA; DT;  
 XX three-dimensional structure; LT; immunoprotective; infection.

XX OS Escherichia coli.

XX PN US5965385-A.

XX PD 12-OCT-1999.

XX PF 06-JUN-1995; 95US-0467974.

XX PR 22-AUG-1994; 94US-0292968.

XX PR 24-AUG-1993; 93US-0110947.

XX PR 31-MAY-1994; 94US-0251121.

XX PA (CONN-) CONNAUGHT LAB LTD.

XX PA (UYAL-) UNIV ALBERTA.

XX PI Read RJ, Stein PE, Cockle SA, Loosmore S, Klein MH, Armstrong GD;  
 XX Hazes B, Oomen RP;

XX DR WPI; 1999-579908/49.

XX PT New method for producing modified pertussis holotoxin -

XX PS Example 3; Fig 5; 41pp; English.

XX CC A method has been developed of producing a modified pertussis holotoxin,  
 CC involving analysis of the 3-dimensional form of the crystalline  
 CC holotoxin. The pertussis holotoxin modification process comprises:  
 CC (1) identification of at least one amino acid (aa) residue of the  
 CC holotoxin for modification by analysing the 3-dimensional form of the  
 CC crystalline holotoxin, in relation to known information of the protein  
 CC structure and function; (2) effecting mutagenesis (by removing or  
 CC replacing a nucleotide sequence encoding at least one (aa) of a tox  
 CC operon; and (3) expressing mutant tox box in a Bordetella organism to  
 CC produce the modified holotoxin. This method is used for modifying  
 CC pertussis holotoxin, by studying its 3-dimensional crystalline  
 CC structure. Modifying the holotoxin, alters its biological properties.  
 CC By analysing the 3-dimensional crystalline structure of the pertussis  
 CC holotoxin, functional (aa) which affect biological properties of the  
 CC pertussis holotoxin can be identified. This can be used to predict (aa)  
 CC which contribute to the toxicity of the holotoxin to produce  
 CC immunoprotective, genetically-detoxified analogues of pertussis  
 CC holotoxin. The present sequence represents an ADP-ribosylating toxin  
 CC B-subunit peptide used in the exemplification of the present  
 CC invention.

SQ Sequence 93 AA;

Query Match 100.0%; Score 39; DB 20; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
 |||||  
 Db 41 evpgsqh 47

RESULT 15

AAW95226  
 ID AAW95226 standard; peptide; 93 AA.

XX

AAW95226;  
 16-MAR-1999 (first entry)  
 E. coli heat-labile toxin (LT) beta-subunit sequence.  
 Pertussis holotoxin; PT; modified; effector; toxicity; cell binding;  
 enzyme; ADP-ribosylation; toxin; crystalline; X-ray crystallography;  
 structural analysis; interacting site; mitogenicity; adjuvanticity;  
 heat-labile; LT.  
 Escherichia coli.  
 US5856122-A.  
 05-JAN-1999.  
 22-AUG-1994; 94US-0292968.  
 22-AUG-1994; 94US-0292968.  
 24-AUG-1993; 93US-0110947.  
 31-MAY-1994; 94US-0251121.  
 (UYAL-) UNIV ALBERTA.  
 Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;  
 Oomen RP, Read RJ, Stein PE;  
 WPI; 1999-105104/09.  
 Modifications to e.g. enzymatic activity, mitogenicity and cell  
 binding of pertussis holotoxin - by identifying interaction sites of  
 a molecule with crystalline toxin and modifying the identified site  
 Example 3; Fig 5; 40pp; English.  
 The invention relates to methods of preparing a pertussis holotoxin (PT)  
 having a modified biological activity. One method comprises identifying  
 at least 1 site in a PT that interacts with a molecule that is capable of  
 forming a complex with the holotoxin and which molecule is an effector  
 molecule which is an adenine nucleotide and which site contributes to  
 toxicity, cell binding or enzymatic activity of PT. The functional  
 interacting site(s) are identified by analysing the three dimensional  
 structure of crystalline PT, determined by X-ray crystallography. The  
 identified interacting site(s) are modified to alter toxicity, cell  
 binding or enzyme activity of the PT. The methods can be used to alter a  
 biological activity such as toxicity, enzymatic activity, mitogenicity,  
 cell binding and adjuvanticity of the PT. The three-dimensional structure  
 of PT have functional and/or structural resemblance to other bacterial  
 toxins such as diphtheria toxin (DT), Pseudomonas exotoxin A (ETA), the  
 heat-labile toxin of E. coli (LT) and verotoxin-1 (VT). The present  
 sequence represents the beta-subunit of LT toxin.  
 Sequence 93 AA;  
 Query Match 100.0%; Score 39; DB 20; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EVPGSQH 7  
 Db 41 evpgsqh 47  
 RESULT 16  
 AAY68365  
 ID AAY68365 standard; Peptide; 93 AA.  
 AC AAY68365;  
 XX  
 XX  
 DT 17-APR-2000 (first entry)  
 XX

DE Heat labile toxin B subunit SEQ ID NO:26.  
 XX Pertussis toxin; heat labile toxin; verotoxin; exotoxin A; holotoxin;  
 KW diphtheria toxin; ADP-ribosylating toxin; mannose binding protein;  
 KW infection; crystal structure; X-ray crystallography; detoxification;  
 KW immunogenic.  
 XX Escherichia coli.  
 OS US6018022-A.  
 PN 25-JAN-2000.  
 XX 06-JUN-1995; 95US-0467976.  
 XX 22-AUG-1994; 94US-0292968.  
 PR 24-AUG-1993; 93US-0110947.  
 PR 31-MAY-1994; 94US-0251121.  
 XX (CONN-) CONNAUGHT LAB LTD.  
 PA (UYAL-) UNIV ALBERTA.  
 XX Read RJ, Stein PE, Cockle SA, Loosmore S, Klein MH, Armstrong GD;  
 PI Hazes B, Oomen RP;  
 PI WPI; 2000-136703/12.  
 DR Crystalline form of isolated pertussis holotoxin useful in studying  
 PT proteins which have functional resemblance -  
 XX Example 3; Fig 5; 42pp; English.  
 CC The present invention describes a crystalline form of isolated  
 CC pertussis holotoxin, in which the molecules of pertussis toxin have  
 CC a three dimensional structure represented in the specification,  
 CC complexed with a polysaccharide molecule capable of forming a complex  
 CC with the holotoxin. The crystalline form of the pertussis holotoxin  
 CC can be used in a comparison with other proteins which have functional  
 CC resemblance to pertussis holotoxin with the aim of modifying other  
 CC proteins. Identifying the unknown sites of toxicity by comparison  
 CC with the three dimensional structure of pertussis holotoxin provides a  
 CC technique for detoxification of toxins to produce useful immunogenic  
 CC but non-toxic analogues. It can also be used as a primary standard to  
 CC measure the quantity, purity or efficacy of less pure compositions  
 CC containing pertussis toxin. AAY68340 to AAY68385 represent peptides  
 CC used in the exemplification of the present invention.  
 XX Sequence 93 AA;  
 SQ  
 Query Match 100.0%; Score 39; DB 21; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EVPGSQH 7  
 Db 41 evpgsqh 47  
 RESULT 17  
 AAB66239  
 ID AAB66239 standard; Protein; 93 AA.  
 XX  
 XX AAB66239;  
 AC  
 XX 03-APR-2001 (first entry)  
 DT  
 DE E coli verotoxin-1 B subunit SEQ ID NO: 26.  
 XX  
 XX Pertussis toxin; crystal structure; whooping cough; biological activity;  
 KW lymphocytosis-promoting factor; histamine-sensitising factor;  
 KW islet-activating protein.  
 XX

OS Escherichia coli.  
 PN US6168928-B1.  
 XX  
 XX 02-JAN-2001.  
 PD  
 XX 21-MAY-1998; 98US-0082514.  
 XX  
 XX 22-AUG-1994; 94US-0292968.  
 PR  
 PR 24-AUG-1993; 93US-0110947.  
 PR  
 PR 31-MAY-1994; 94US-0251121.  
 XX  
 XX (CONN-) CONNAUGHT LAB LTD.  
 PA  
 XX Read RJ, Cockle SA, Oomen RP, Loosmore S, Klein MH, Armstrong GD;  
 PI Hazes B, Stein PE;  
 PI  
 XX WPI; 2001-122260/13.  
 DR  
 XX  
 XX Modifying pertussis holotoxin to produce detoxified PT analogs,  
 PT comprising analyzing crystalline structure of toxin, to identify sites  
 PT of toxicity, cell binding or enzyme activity of PT and modifying  
 PT identified site -  
 XX  
 XX Example 3; Fig 5; 41pp; English.  
 PS  
 XX The present invention provides a method for producing a pertussis toxin  
 CC (also designated lymphocytosis-promoting factor, histamine-sensitising  
 CC factor and islet activating protein) with a modified biological activity,  
 CC involving analysing the crystal structure of the protein to identify  
 CC active sites which can then be modified. This may lead to an alteration  
 CC in the toxicity, cell binding or enzyme activity of the toxin. This can  
 CC be used in the production of immunoprotective analogues of pertussis  
 CC toxin. Pertussis toxin is the cause of whooping cough following infection  
 CC by Bordetella pertussis.  
 XX  
 XX Sequence 93 AA;  
 SQ

Query Match 100.0%; Score 39; DB 22; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
 DB 41 evpgsqh 47  
 |||||

RESULT 18  
 AAP50340  
 ID AAP50340 standard; protein; 103 AA.  
 XX  
 XX AAP50340;  
 AC  
 XX 01-DEC-1991 (first entry)  
 DT  
 XX Sequence of sub-unit B of cholera toxin.  
 DE  
 XX Vaccine; cholera; heat-labile E.coli toxin.  
 KW  
 XX Vibrio cholera.  
 OS  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Region 50..64  
 FT /note= "claimed"  
 FT Region 8..20  
 FT /note= "claimed"  
 FT Region 45..64  
 FT /note= "claimed"  
 FT  
 XX DE3430894-A.  
 PN  
 XX 14-MAR-1985.  
 PD

XX 22-AUG-1984; 84DE-3430894.  
 PF  
 XX 23-AUG-1983; 83IL-0069558.  
 PR  
 XX (YEDA ) YEDA RES & DEV CO LTD.  
 PA  
 XX Sela M, Arnon R, Jacob CO;  
 PI  
 XX WPI; 1985-069683/12.  
 DR  
 XX Vaccines against cholera and heat-labile E. coli toxin - contg.  
 PT cholera toxin fragment coupled to carrier  
 PT  
 XX Example; Fig 1; 24pp; German.  
 PS  
 XX The inventors claim vaccines against cholera and heat-labile E.coli  
 CC toxin contg. cholera toxin fragment coupled to carrier. The toxin is  
 CC esp. the fragments defined in FT, above.  
 CC  
 XX Sequence 103 AA;  
 SQ

Query Match 100.0%; Score 39; DB 6; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
 DB 51 evpgsqh 57  
 |||||

RESULT 19  
 AAW04857  
 ID AAW04857 standard; Protein; 103 AA.  
 XX  
 XX AAW04857;  
 AC  
 XX 21-FEB-1997 (first entry)  
 DT  
 XX Synthetic cholera toxin B subunit.  
 DE  
 XX Bordetella pertussis; whooping cough; recombinant construct;  
 KW cholera toxin B subunit; enzyme; antigen; immunogen; allergen;  
 KW enzyme inhibitor; hormone; lymphokine; immunoglobulin; toxin;  
 KW structural protein; receptor; heterologous gene; leader; promoter.  
 XX  
 XX Synthetic.  
 OS  
 XX WO9626282-A1.  
 PN  
 XX 29-AUG-1996.  
 PD  
 XX 23-FEB-1996; 96WO-CA00107.  
 PF  
 XX 23-FEB-1995; 95US-0393334.  
 PR  
 XX (CONN-) CONNAUGHT LAB LTD.  
 PA  
 XX Klein MH, Loosmore SM, Yacoob RK, Zealey GR;  
 PI  
 XX WPI; 1996-425088/42.  
 DR  
 DR N-PSDB; AAP38038.  
 DR  
 XX Recombinant constructs for expressing and opt. secreting proteins in  
 PT Bordetella - comprise Bordetella promoter coupled to non-Bordetella,  
 PT esp. cholera B toxin, gene or coupled to non-Bordetella leader and  
 PT gene of interest  
 PT  
 XX Example 1; Figure 1; 61pp; English.  
 PS  
 XX Recombinant constructs comprising a promoter functional in  
 CC Bordetella operatively linked to a heterologous gene or a non-  
 CC

CC Bordetella leader sequence for secretion of a gene product which may  
 CC or may not be of Bordetella origin, can be used for the expression  
 CC in Bordetella of enzymes, antigens, immunogens, allergens, enzyme  
 CC inhibitors, hormones, lymphokines, immunoglobulins or their  
 CC fragments, toxins, mammalian proteins, structural proteins or  
 CC receptors. The Bordetella strains are particularly engineered to  
 CC express the cholera toxin B subunit (this sequence). The promoters  
 CC used in the constructs are selected from the Bordetella pertussis  
 CC tox, fla promoters or the high molecular weight (hmw) outer membrane  
 CC promoter of non typable Haemophilus influenzae; leaders used in the  
 CC constructs are selected from the cholera toxin B leader (CTB-L), the  
 CC pertussis toxin subunit S1 leader (S1-L) and the pertussin pertactin  
 CC leader (PRN-L); and genes used in the constructs are selected from a  
 CC novel synthetic cholera toxin B gene (ctb) and the hmw1 and hmw2  
 CC genes of Haemophilus influenzae.

XX  
 SQ Sequence 103 AA;

Query Match 100.0%; Score 39; DB 17; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 1.4; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0;

Oy 1 EYPGSOH 7  
 Db 51 evpgsqh 57  
 |||||

## RESULT 20

AAW06606  
 ID AAR94939 standard; Protein; 103 AA.

XX  
 AC AAR94939;

XX  
 DT 31-OCT-1996 (first entry)

XX Heat labile enterotoxin B subunit (LT-B) E.coli.

XX Toxin; subunit; vaccine; transgenic plant; immunogen; antigen;  
 KW adjuvant; immunisation.

XX Escherichia coli.

XX WO9612801-A1.

XX 02-MAY-1996.

XX 24-OCT-1995; 95WO-US13376.

XX 24-OCT-1994; 94US-0328716.

XX (TULA ) TULANE EDUCATIONAL FUND.  
 PA (TEXA ) UNIV TEXAS A & M SYSTEM.

XX Arntzen CJ, Clements JD, Haq TA, Mason HS;

XX WPI; 1996-230602/23

XX N-PSDB; AAT18799, AAT18800.

XX Transgenic plants contg. E. coli heat labile enterotoxin subunits -  
 PT used as oral vaccines for animals which consume the plant

XX Disclosure; Page 100-101; 130pp; English.

XX A transgenic plant comprising or expressing a DNA sequence encoding  
 CC an immunogenic agent can be used as an oral vaccine for animals.  
 CC The vaccine is administered by the oral consumption of the plant and  
 CC provides the first known functional method for immunising animals  
 CC using transgenic plants, where the plants express bacterial antigens  
 CC that act as both immunogens and adjuvants. The method provides an  
 CC inexpensive production and delivery system for such antigens to  
 CC animals. This is the LT-B Escherichia coli toxin subunit and its  
 CC coding sequence was used in the construction of such a transgenic

CC plant. The immunogenic agent preferably comprises the LT-B or CT-B  
 CC (cholera toxin B subunit) or optionally LT-A or CT-A.

XX SQ Sequence 103 AA;

Query Match 100.0%; Score 39; DB 17; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 1.4; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0;

Oy 1 EYPGSOH 7  
 Db 51 evpgsqh 57  
 |||||

## RESULT 21

AAW06606  
 ID AAW06606 standard; Protein; 103 AA.

XX  
 AC AAW06606;

XX  
 DT 06-AUG-1997 (first entry)

XX Cholera toxin B subunit/heat labile enterotoxin B subunit hybrid.

XX Heat labile; tolerant; stable; hybrid; cholera toxin B subunit; CTB;  
 KW heat labile enterotoxin B subunit; LTb; vaccine; immunisation;  
 KW immunogen; antigen; broad spectrum; cross reactive; diarrhoea;  
 KW vomiting; food poisoning.

XX Chimeric - Vibrio cholerae.

OS Chimeric - Enterotoxigenic Escherichia Coli.

XX Key Location/Qualifiers  
 FT Misc-difference 1

FT /label= substitution  
 FT /note= "wild-type Thr replaced by Ala"

FT Misc-difference 94

FT /label= substitution  
 FT /note= "wild-type His replaced by Asn"

FT Misc-difference 95

FT /label= substitution  
 FT /note= "wild-type Ala replaced by Ser"

XX WO9634893-A1.

XX 07-NOV-1996.

XX 02-MAY-1996; 96WO-SE00570.

XX 05-MAY-1995; 95SE-0001682.

XX (HOLM/) HOLMGREN J.  
 PA (LEBE/) LEBENS M R.

XX Holmgren J, Lebens MR;

XX WPI; 1996-506108/50.

XX N-PSDB; AAT45576.

XX Heat-labile enterotoxin B sub-unit and cholera toxin B sub-unit  
 PT hybrid protein - opt. fused to immunogenic sequence for use in  
 PT vaccines against enterotoxin-induced illness

XX Claim 3; Page -; 32pp; English.

XX AAW06606 is a mature cholera toxin B subunit (CTB)/heat labile  
 CC enterotoxin B subunit (LTb) hybrid molecule. In mature CTB protein,  
 CC certain amino acids (aa) were replaced with corresponding aa from  
 CC heat-labile enterotoxin B subunit (LTb). The specific amino acid  
 CC substitutions impart LTb-specific epitope characteristics to  
 CC immunogenic mature CTB. The hybrid molecules have increased  
 CC cross-reactivity and are suitable for a broad spectrum vaccine to



CC protect against enterotoxigenic illness. Immunogenic proteins  
 CC comprising the hybrid molecules can be used to treat, or in a  
 CC vaccine to prevent, enterotoxin-induced illness, e.g. diarrhoea  
 CC and vomiting, in humans and animals.  
 CC Note - This sequence does not appear in the specification, it is  
 CC a claimed mutant sequence of mature cholera toxin B subunit (see  
 CC AAW06605).  
 XX  
 SQ Sequence 103 AA;

Query Match 100.0%; Score 39; DB 17; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
 |||||  
 Db 51 evpgsqh 57

RESULT 22  
 AAW06607  
 ID AAW06607 standard; Protein; 103 AA.  
 XX  
 AC AAW06607;  
 DT 06-AUG-1997 (first entry)  
 XX  
 DE Cholera toxin B subunit/heat labile enterotoxin B subunit hybrid.  
 XX  
 KW Heat labile; tolerant; stable; hybrid; cholera toxin B subunit; CTB;  
 KW heat labile enterotoxin B subunit; LTB; vaccine; immunisation;  
 KW immunogen; antigen; broad spectrum; cross reactive; diarrhoea;  
 KW vomiting; food poisoning.  
 XX  
 OS Chimeric - Vibrio cholerae.  
 OS Chimeric - Enterotoxigenic Escherichia Coli.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1..25  
 FT /label= substitution  
 FT /note= "the first 25 amino acids of mature  
 FT wild-type cholera toxin B subunit are  
 FT replaced with the first 25 amino acids  
 FT of mature enterotoxin B subunit"  
 XX  
 PN W09634893-A1.  
 XX  
 PD 07-NOV-1996.  
 XX  
 PF 02-MAY-1996; 96WO-SE00570.  
 XX  
 PR 05-MAY-1995; 95SE-0001682.  
 XX  
 PA (HOLM/) HOLMGREN J.  
 PA (LEBE/) LEBENS M R.  
 XX  
 PI Holmgren J, Lebens MR;  
 XX  
 DR WPI: 1996-506108/50.  
 DR N-PSDB; AAT43577.  
 XX  
 XX Heat-labile enterotoxin B sub-unit and cholera toxin B sub-unit  
 PT hybrid protein - opt. fused to immunogenic sequence for use in  
 PT vaccines against enterotoxin-induced illness  
 XX  
 PS Claim 4; Page -; 32pp; English.  
 XX  
 CC AAW06607 is a mature cholera toxin B subunit (CTB)/heat labile  
 CC enterotoxin B subunit (LTB) hybrid molecule. In mature CTB protein,  
 CC certain amino acids (aa) were replaced with corresponding aa from  
 CC heat-labile enterotoxin B subunit (LTB). The specific amino acid  
 CC substitutions impart LTB-specific epitope characteristics to

CC immunogenic mature CTB. The hybrid molecules have increased  
 CC cross-reactivity and are suitable for a broad spectrum vaccine to  
 CC protect against enterotoxigenic illness. Immunogenic proteins  
 CC comprising the hybrid molecules can be used to treat, or in a  
 CC vaccine to prevent, enterotoxin-induced illness, e.g. diarrhoea  
 CC and vomiting, in humans and animals.  
 CC Note - This sequence does not appear in the specification, it is  
 CC a claimed mutant sequence of mature cholera toxin B subunit (see  
 CC AAW06605).  
 XX  
 SQ Sequence 103 AA;

Query Match 100.0%; Score 39; DB 17; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
 |||||  
 Db 51 evpgsqh 57

RESULT 23  
 AAW80808  
 ID AAW80808 standard; Protein; 103 AA.  
 XX  
 AC AAW80808;  
 DT 29-JAN-1999 (first entry)  
 XX  
 DE Amino acid sequence of the wild type cholera toxin B subunit.  
 XX  
 KW Cholera toxin B subunit; nontoxic subunit; adjuvant; coadministration;  
 KW antigen; bird; animal; mucosal; vaccine.  
 XX  
 OS Vibrio cholerae.  
 XX  
 PN W09845324-A1.  
 XX  
 PD 15-OCT-1998.  
 XX  
 PF 03-APR-1998; 98WO-US06725.  
 XX  
 PR 04-APR-1997; 97US-0043410.  
 XX  
 PA (KIYO/) KIYONO H.  
 PA (MCGH/) MCGHEE J R.  
 PA (TAKE/) TAKEDA Y.  
 PA (UABR-) UAB RES FOUND.  
 PA (YAMA/) YAMAMOTO S.  
 XX  
 PI Kiyono H, Mcghee JR, Takeda Y, Yamamoto S;  
 XX  
 DR WPI: 1998-594478/50.  
 XX  
 PT New mutant cholera toxin selected from a group comprising nontoxic  
 PT subunits/derivatives - effective as an adjuvant when coadministered  
 PT with an antigen to birds and mammals  
 XX  
 PS Disclosure; Fig 1B; 43pp; English.  
 XX  
 CC This is the amino acid sequence of the cholera toxin B subunit used in  
 CC the method of the invention involving the use of nontoxic subunits as  
 CC an effective adjuvant in coadministration of an antigen to birds  
 CC and animals. In addition to the use of the toxin as an mucosal  
 CC adjuvant, it also provides a vaccine comprising the toxin, an  
 CC immunogenic amount of an antigen, and a pharmaceutically acceptable  
 CC carrier. The toxin can be used with single/multiple vaccines, and it  
 CC enables the possibility for commercial mucosal adjuvants for use in  
 CC humans, since these are more effective and safer than vaccines  
 CC administered subcutaneously.  
 XX  
 SQ Sequence 103 AA;

Query Match 100.0%; Score 39; DB 19; Length 103;  
Best Local Similarity 100.0%; Pred. No. 1.4; 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0;

QY 1 EVPGSQH 7  
|||||  
DB 51 evpgsqh 57

## RESULT 24

AA04163  
ID AAR04163 standard; protein; 118 AA.

XX AC AAR04163;

XX DT 10-SEP-1990 (first entry)

XX DE Cholera Toxin B-subunit.

XX KW cholera toxin B-subunit; hybrid protein; heterologous IgA active antigen.

XX OS synthetic.

XX FH Key Location/Qualifiers

FT misc\_difference 18..18

FT /\*label= His or Tyr

FT region 1..11

FT /\*label= signal peptide

FT /\*note= absent from mature protein

XX PN W09003437-A.

XX PD 05-APR-1990.

XX PF 27-SEP-1989; 89WO-0000495.

XX PR 27-SEP-1988; 88FR-0012627.

XX PA (UYLI-) L'UNIVERSITE DE L'ETAT A LIEGE.

XX PI L'Hoir C, Renard A, Martial J;

XX DR WPI; 1990-132273/17.

XX DR N-NSDB; Q04046.

XX PT New hybrid protein, useful in vaccines -

PT contains cholera toxin b subunit and heterologous IgA active

PT antigenic sequence.

XX PS Disclosure; ; pp; French.

XX CC Mature cholera toxin B-subunit is obtained when the signal peptide is  
cleaved off. There is an Ochre codon at position 343-5; the sequence  
downstream from it is part of a plasmid.

XX SQ Sequence 118 AA;

Query Match 100.0%; Score 39; DB 11; Length 118;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
|||||  
DB 62 evpgsqh 68

## RESULT 25

AA04163  
ID AAR04163 standard; protein; 124 AA.

XX

AC AAP93561;  
XX 06-JUN-1990 (first entry)  
DI B subunit of the heat-labile enterotoxin (LT-B) derived from E. coli.  
XX B subunit; heat-labile enterotoxin; LT-B; Escherichia coli; malaria;  
DE circumsporozoite protein; fusion protein; live recombinant vaccine;  
XX Salmonella; epitope.  
KW Escherichia coli.  
XX OS  
XX FH Key Location/Qualifiers  
FT Peptide 1..22  
FT /note="Signal peptide"  
FT Protein 23..124  
FT /note="Mature LT-B"

XX PN W08902924-A.

XX PD 06-APR-1989.

XX PF 30-SEP-1988; 88WO-US03376.

XX PR 02-OCT-1987; 87US-0104735.

XX PA (PRAX-) PRAXIS BIOLOGICS IN.

XX PI Brey RN, Majarian WR, Pillai S, Hockmeyer WT;

XX DR WPI; 1989-114399/15.

XX DR N-PSDB; AAN90747.

XX PT Live recombinant vaccine for malaria -  
comprising attenuated entero-invasive bacterium contg. DNA  
encoding epitope of malaria parasite

XX PS Fig 3; p. 3/17; 105pp; English.

XX CC In the patent, the DNA encoding LT-B is expressed as part of a fusion  
protein with an epitope of a malaria parasite, eg Region I or Region II  
or a repeat region of circumsporozoite protein antigen (CS) (AAP93560)  
from Plasmodium berghei. Pref. the fusion gene is inserted into  
attenuated Salmonella enteritidis under the left promoter control of  
lambda. Such bacteria can multiply in the host without causing disease or  
disorder and express CS that will induce a protective immune response  
against malaria and can be used in vaccines. Such vaccines can be  
multivalent.

XX SQ Sequence 124 AA;

Query Match 100.0%; Score 39; DB 10; Length 124;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
|||||  
DB 72 evpgsqh 78

Search completed: July 16, 2001, 16:35:41  
Job time: 205 sec



***This Page Blank (uspto)***

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 16:32:51 ; Search time 30.3 Seconds  
(without alignments)  
4.654 Million cell updates/sec

Title: US-09-786-648-2

Perfect score: 39

Sequence: 1 EVPSQH 7

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 2014635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:\*
- 1: /cgn2.6/ptodata/2/1aa/5A-COMB.pep.\*
  - 2: /cgn2.6/ptodata/2/1aa/5B-COMB.pep.\*
  - 3: /cgn2.6/ptodata/2/1aa/6A-COMB.pep.\*
  - 4: /cgn2.6/ptodata/2/1aa/6B-COMB.pep.\*
  - 5: /cgn2.6/ptodata/2/1aa/PTCUS-COMB.pep.\*
  - 6: /cgn2.6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	100.0	93	2	US-08-292-968-26
2	39	100.0	93	2	US-08-467-974-26
3	39	100.0	93	2	US-08-467-536-26
4	39	100.0	93	3	US-08-467-976-26
5	39	100.0	93	4	US-09-082-514-26
6	39	100.0	102	3	US-08-952-337-5
7	39	100.0	102	3	US-08-952-337-6
8	39	100.0	103	2	US-08-472-171-2
9	39	100.0	103	2	US-08-894-526-2
10	39	100.0	103	2	US-09-013-047-2
11	39	100.0	103	4	US-09-374-597-2
12	39	100.0	103	4	US-09-191-852-21
13	39	100.0	103	5	PCT-US95-13376-21
14	39	100.0	123	3	US-08-952-337-1
15	39	100.0	123	3	US-08-952-337-2
16	39	100.0	124	2	US-08-747-410-2
17	39	100.0	371	2	US-08-829-026A-6
18	33	84.6	124	1	US-08-449-045C-4
19	33	84.6	124	2	US-08-435-605A-12
20	33	84.6	124	6	5223610-3
21	31	79.5	414	1	US-08-255-471-9
22	31	79.5	459	6	5194375-6
23	30	76.9	262	6	5194375-4
24	30	76.9	329	2	US-08-781-802-8
25	30	76.9	329	4	US-08-694-078-8
26	30	76.9	329	4	US-09-058-260-8
27	30	76.9	459	6	5194375-2

28	30	76.9	775	2	US-08-714-070A-1	Sequence 1, Appli
29	30	76.9	805	1	US-08-045-806-2	Sequence 2, Appli
30	30	76.9	805	1	US-08-366-051B-2	Sequence 2, Appli
31	30	76.9	855	2	US-09-027-337-2	Sequence 2, Appli
32	30	76.9	1285	2	US-08-540-406-6	Sequence 6, Appli
33	30	76.9	1285	3	US-08-656-055-6	Sequence 6, Appli
34	30	76.9	1285	4	US-08-954-668-6	Sequence 6, Appli
35	30	76.9	1285	5	PCT-US95-13233-6	Sequence 6, Appli
36	30	76.9	1299	4	US-08-460-900C-62	Sequence 62, Appli
37	30	76.9	1810	5	PCT-US95-11684-4	Sequence 4, Appli
38	29	74.4	346	2	US-08-602-359A-34	Sequence 34, Appli
39	29	74.4	393	2	US-08-990-379-4	Sequence 4, Appli
40	29	74.4	1036	3	US-08-968-752B-4	Sequence 4, Appli
41	29	74.4	1074	2	US-08-768-147B-2	Sequence 2, Appli
42	29	74.4	1074	3	US-08-968-752B-2	Sequence 2, Appli
43	28	71.8	237	1	US-08-910-973-13	Sequence 13, Appli
44	28	71.8	433	1	US-07-661-610C-2	Sequence 2, Appli
45	28	71.8	468	1	US-08-164-614A-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1  
US-08-292-968-26  
; Sequence 26, Application US/08292968  
; Patent No. 5856122  
; GENERAL INFORMATION:  
; APPLICANT: READ, Randy J.  
; APPLICANT: STEIN, Penelope E.  
; APPLICANT: COCKLE, Stephen A.  
; APPLICANT: OOMEN, Raymond P.  
; APPLICANT: LOOSMORE, Sheena  
; APPLICANT: KLEIN, Michel H.  
; APPLICANT: ARMSTRONG, Glen D.  
; APPLICANT: HAZES, Bart  
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/292,968  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/110,947  
; FILING DATE: 24-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/251,121  
; FILING DATE: 31-MAY-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-388  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 93 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-08-292-968-26

Query Match 100.0%; Score 39; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
|||||  
DB 41 EVPGSQH 47

RESULT 2

US-08-467-974-26  
; Sequence 26, Application US/08467974  
; Patent No. 5965385  
; GENERAL INFORMATION:  
; APPLICANT: READ, Randy J.  
; APPLICANT: STEIN, Penelope E.  
; APPLICANT: COCKLE, Stephen A.  
; APPLICANT: COOMEN, Raymond P.  
; APPLICANT: LOOSMORE, Sheena  
; APPLICANT: KLEIN, Michel H.  
; APPLICANT: ARMSTRONG, Glen D.  
; APPLICANT: HAZES, Bart  
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Slim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467,974  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/467,536  
; FILING DATE: 06-JUN-1995  
; APPLICATION NUMBER: US 08/292,968  
; FILING DATE: 22-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/251,121  
; FILING DATE: 31-MAY-1994  
; APPLICATION NUMBER: US 08/110,947  
; FILING DATE: 24-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-454 MIS:vg  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 93 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-08-467-974-26

Query Match 100.0%; Score 39; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
|||||  
DB 41 EVPGSQH 47

RESULT 3

US-08-467-536-26  
; Sequence 26, Application US/08467536  
; Patent No. 5977304  
; GENERAL INFORMATION:  
; APPLICANT: READ, Randy J.  
; APPLICANT: STEIN, Penelope E.  
; APPLICANT: COCKLE, Stephen A.  
; APPLICANT: COOMEN, Raymond P.  
; APPLICANT: LOOSMORE, Sheena  
; APPLICANT: KLEIN, Michel H.  
; APPLICANT: ARMSTRONG, Glen D.  
; APPLICANT: HAZES, Bart  
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Slim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467,536  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/292,968  
; FILING DATE: 22-AUG-1994  
; APPLICATION NUMBER: US 08/251,121  
; FILING DATE: 31-MAY-1994  
; APPLICATION NUMBER: US 08/110,947  
; FILING DATE: 24-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-455 MIS:vg  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 93 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-467-536-26

Query Match 100.0%; Score 39; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
|||||  
DB 41 EVPGSQH 47

RESULT 4

US-08-467-976-26

```

; Sequence 26, Application US/08467976
; Patent No. 6018022
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J.
; APPLICANT: STEIN, Penelope E.
; APPLICANT: COCKLE, Stephen A.
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: KLEIN, Michel H.
; APPLICANT: ARMSTRONG, Glen D.
; APPLICANT: HAZES, Bart
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,976
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,968
; FILING DATE: 22-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/251,121
; FILING DATE: 31-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/110,947
; FILING DATE: 24-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-453 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-467-976-26

```

```

Query Match 100.0%; Score 39; DB 3; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 EVPGSQH 7
Db 41 EVPGSQH 47

```

```

RESULT 5
US-09-082-514-26
; Sequence 26, Application US/09082514
; Patent No. 6168928
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J.
; APPLICANT: STEIN, Penelope E.
; APPLICANT: COCKLE, Stephen A.
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.

```

```

; APPLICANT: ARMSTRONG, Glen D.
; APPLICANT: HAZES, Bart
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/082,514
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,968
; FILING DATE: 24-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-810
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-082-514-26

Query Match 100.0%; Score 39; DB 4; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7
Db 41 EVPGSQH 47

RESULT 6
US-08-952-337-5
; Sequence 5, Application US/08952337
; Patent No. 6019973
; GENERAL INFORMATION:
; APPLICANT: Holmgren, Jan
; APPLICANT: Lebens, Michael R.
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
; TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS
; FILE REFERENCE: 3846/0D758
; CURRENT APPLICATION NUMBER: US/08/952,337
; CURRENT FILING DATE: 1998-01-05
; EARLIER APPLICATION NUMBER: PCT/SE96/00570
; EARLIER FILING DATE: 1996-05-02
; EARLIER APPLICATION NUMBER: SE 9501682-0
; EARLIER FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows version 3.0
; SEQ ID NO 5
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Vibrio cholerae
; US-08-952-337-5

```

Query Match 100.0%; Score 39; DB 3; Length 102;  
Best Local Similarity 100.0%; Pred. No. 0.62;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
| | | | | | |  
Db 50 EYPGSOH 56

## RESULT 7

US-08-952-337-6  
; Sequence 6, Application US/08952337  
; Patent No. 6019973  
; GENERAL INFORMATION:  
; APPLICANT: Holmgren, Jan  
; APPLICANT: Lebens, Michael R.  
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE  
; FILE REFERENCE: 3846/00758  
; CURRENT APPLICATION NUMBER: US/08/952.337  
; EARLIER FILING DATE: 1998-01-05  
; EARLIER APPLICATION NUMBER: PCT/SE96/00570  
; EARLIER FILING DATE: 1996-05-02  
; EARLIER APPLICATION NUMBER: SE 9501682-0  
; EARLIER FILING DATE: 1995-05-05  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 102  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-08-952-337-6

Query Match 100.0%; Score 39; DB 3; Length 102;  
Best Local Similarity 100.0%; Pred. No. 0.62;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
| | | | | | |  
Db 50 EYPGSOH 56

## RESULT 8

US-08-472-171-2  
; Sequence 2, Application US/08472171  
; Patent No. 5932714  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yacoub, Reza K.  
; APPLICANT: Zealey, Gavin R.  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Expression of Gene Products From  
; TITLE OF INVENTION: Genetically Manipulated Strains of Bordetella  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 330 University Avenue, Suite 701  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/472.171  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/393,334

; FILING DATE: 23-FEB-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-507 MIS:vg  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 416-595-1155  
; TELEFAX: 416-595-1163  
; TELEX: 065-24567 Simbas  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 103 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-472-171-2

Query Match 100.0%; Score 39; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.62;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
| | | | | | |  
Db 51 EYPGSOH 57

## RESULT 9

US-08-894-526-2  
; Sequence 2, Application US/08894526  
; Patent No. 5942418  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M  
; APPLICANT: Yacoub, Reza K  
; APPLICANT: Zealey, Gavin R  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: EXPRESSION OF GENE PRODUCTS FROM  
; TITLE OF INVENTION: GENETICALLY MANIPULATED STRAINS OF BORDETELLA  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/894,526  
; FILING DATE: 01-DEC-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-724 MIS:jb  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 103 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-894-526-2

Query Match 100.0%; Score 39; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.62;



Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
|||||||  
Db 51 EVPGSQH 57

## RESULT 10

US-09-013-047-2  
; Sequence 2, Application US/09013047  
; Patent No. 5998168  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yacoub, Reza K.  
; APPLICANT: Zealey, Gavin R.  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Expression Of Gene Products From  
; TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 330 University Avenue, 6th Floor  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/013,047  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/472,171  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/393,334  
; FILING DATE: 23-FEB-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-507 MIS:vg  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 416-595-1155  
; TELEFAX: 416-595-1163  
; TELEX: 065-24567 Simbas  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 103 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-013-047-2

Query Match 100.0%; Score 39; DB 2; Length 103;

Best Local Similarity 100.0%; Pred. No. 0.62;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
|||||||  
Db 51 EVPGSQH 57

## RESULT 11

US-09-374-597-2  
; Sequence 2, Application US/09374597  
; Patent No. 6140082  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.

; APPLICANT: Yacoub, Reza K.  
; APPLICANT: Zealey, Gavin R.  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Expression Of Gene Products From  
; TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 330 University Avenue, 6th Floor  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/374,597  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/393,334  
; FILING DATE: FEBRUARY 23, 1995  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-964  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 416-595-1155  
; TELEFAX: 416-595-1163  
; TELEX: 065-24567 Simbas  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 103 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-374-597-2

Query Match 100.0%; Score 39; DB 4; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.62;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
|||||||  
Db 51 EVPGSQH 57

RESULT 12  
US-09-191-852-21  
; Sequence 21, Application US/09191852  
; Patent No. 6194560  
; GENERAL INFORMATION:  
; APPLICANT: Charles J. Arntzen, Hugh S. Mason, and Tariq A. Haq  
; TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fulbright & Jaworski L.L.P.  
; STREET: 1301 McKinney, Suite 5100  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77010  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/191.852  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA: PCT/US95/13376  
APPLICATION NUMBER: PCT/US95/13376  
FILING DATE: 24-OCT-1995  
APPLICATION NUMBER: 08/817.906  
FILING DATE: 04-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Fox, David L.  
REGISTRATION NUMBER: 40,612  
REFERENCE/DOCKET NUMBER: P01590US1  
TELEPHONE: 713-651-5151  
TELEFAX: 713-651-5246  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-09-191-852-21

Query Match 100.0%; Score 39; DB 4; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.62;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
Db 51 EVPGSQH 57

RESULT 13  
PCT-US95-13376-21  
Sequence 21, Application PC/TUS9513376  
GENERAL INFORMATION:  
APPLICANT: The Texas A&M University System  
APPLICANT: 310 Wisenbaker  
APPLICANT: College Station, Texas 77843-3369  
TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pravel, Hewitt, Kimball & Krieger  
STREET: 1177 West Loop South, 10th Floor  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77027-9095  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13376  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/328,716  
FILING DATE: 24-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jones, John W.  
REGISTRATION NUMBER: 31,380  
REFERENCE/DOCKET NUMBER: 36170/3P  
TELEPHONE: 713-850-0909  
TELEFAX: 713-850-0165  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
PCT-US95-13376-21

Query Match 100.0%; Score 39; DB 5; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.62;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
Db 51 EVPGSQH 57

RESULT 14  
US-08-952-337-1  
Sequence 1, Application US/08952337  
Patent No. 6019973  
GENERAL INFORMATION:  
APPLICANT: Holmgren, Jan  
APPLICANT: Lebens, Michael R.  
TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE  
FILE REFERENCE: 3846/00758  
CURRENT APPLICATION NUMBER: US/08/952,337  
CURRENT FILING DATE: 1998-01-05  
EARLIER APPLICATION NUMBER: PCT/SE96/00570  
EARLIER FILING DATE: 1996-05-02  
EARLIER APPLICATION NUMBER: SE 9501682-0  
EARLIER FILING DATE: 1995-05-05  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 123  
TYPE: PRT  
ORGANISM: Vibrio cholerae  
US-08-952-337-1

Query Match 100.0%; Score 39; DB 3; Length 123;  
Best Local Similarity 100.0%; Pred. No. 0.75;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
Db 71 EVPGSQH 77

RESULT 15  
US-08-952-337-2  
Sequence 2, Application US/08952337  
Patent No. 6019973  
GENERAL INFORMATION:  
APPLICANT: Holmgren, Jan  
APPLICANT: Lebens, Michael R.  
TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE  
FILE REFERENCE: 3846/00758  
CURRENT APPLICATION NUMBER: US/08/952,337  
CURRENT FILING DATE: 1998-01-05  
EARLIER APPLICATION NUMBER: PCT/SE96/00570  
EARLIER FILING DATE: 1996-05-02  
EARLIER APPLICATION NUMBER: SE 9501682-0  
EARLIER FILING DATE: 1995-05-05  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 123  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-08-952-337-2

Query Match 100.0%; Score 39; DB 3; Length 123;  
Best Local Similarity 100.0%; Pred. No. 0.75;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
|||||||  
Db 71 EVPGSQH 77

RESULT 16  
US-08-747-410-2  
; Sequence 2, Application US/08747410  
; Patent No. 5993820  
; GENERAL INFORMATION:  
; APPLICANT: BAGDASARIAN, Michael  
; APPLICANT: IRLAND, James  
; TITLE OF INVENTION: CHIMERIC LTB VACCINES  
; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt  
; STREET: 3100 No. 5993820west Center, 90 South Seventh St  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/747,410  
; FILING DATE: 12-NOV-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:

; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kettelberger, Denise M  
; REGISTRATION NUMBER: 33,924  
; REFERENCE/DOCKET NUMBER: 11526.1-US-01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612/371-5268  
; TELEFAX: 612/332-9081  
; TELEX:

; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 124 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: Internal  
; ORIGINAL SOURCE:  
US-08-747-410-2

Query Match 100.0%; Score 39; DB 2; Length 124;  
Best Local Similarity 100.0%; Pred. No. 0.76;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
|||||||  
Db 72 EVPGSQH 78

RESULT 17  
US-08-829-026A-6  
; Sequence 6, Application US/08829026A  
; Patent No. 5837825  
; GENERAL INFORMATION:  
; APPLICANT: Meinersmann, Richard J.  
; APPLICANT: Khoury, Christian A.  
; TITLE OF INVENTION: Campylobacter Jejuni Flagellin-Escherichia Coli LT-B Fusion P

; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Janelle S. Graeter  
; STREET: Room 411, Building 005, BARC-W  
; CITY: Beltsville  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20705  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/829,026A  
; FILING DATE: 18-AUG-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Graeter, Janelle S.  
; REGISTRATION NUMBER: 35,024  
; REFERENCE/DOCKET NUMBER: 0106.97  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 301-504-5676  
; TELEFAX: 301-504-5060

; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 371 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-829-026A-6

Query Match 100.0%; Score 39; DB 2; Length 371;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
|||||||  
Db 55 EVPGSQH 61

RESULT 18  
US-08-449-045C-4  
; Sequence 4, Application US/08449045C  
; Patent No. 5770203  
; GENERAL INFORMATION:  
; APPLICANT: Burnette, Neal W.  
; APPLICANT: Kaslow, Harvey R.  
; TITLE OF INVENTION: RECOMBINANT DNA-DERIVED CHOLERA TOXIN  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Inc.  
; STREET: 1840 De Havilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/449,045C  
; FILING DATE: 24-MAY-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/694,733  
; FILING DATE: 02-MAY-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/271,222

;; FILING DATE: 06-JUL-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Mazza, Richard J.  
;; REGISTRATION NUMBER: 27,657  
;; REFERENCE/DOCKET NUMBER: A-196C  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 124 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-449-045C-4

Query Match 84.6%; Score 33; DB 1; Length 124;  
Best Local Similarity 85.7%; Pred. No. 11;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
||| |||  
Db 72 EVPSSQH 78

RESULT 19  
US-08-435-605A-12  
; Sequence 12, Application US/08435605A  
; Patent No. 5874287  
; GENERAL INFORMATION:  
; APPLICANT: Burnette, W. Neal  
; APPLICANT: Kaslow, Harvey R.  
; TITLE OF INVENTION: RECOMBINANT DNA-DERIVED CHOLERA TOXIN  
; TITLE OF INVENTION: SUBUNIT ANALOGS  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Inc.  
; STREET: 1840 De Havilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/435,605A  
; FILING DATE: 05-MAY-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mazza, Richard J.  
; REGISTRATION NUMBER: 27,657  
; REFERENCE/DOCKET NUMBER: A-196B  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 124 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-435-605A-12

Query Match 84.6%; Score 33; DB 2; Length 124;  
Best Local Similarity 85.7%; Pred. No. 11;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
||| |||  
Db 72 EVPSSQH 78

RESULT 20

5223610-3  
; Patent No. 5223610  
; APPLICANT: Burton, Frank H.; Sutcliffe, Gregor  
; TITLE OF INVENTION: CHOLERA TOXIN GENE REGULATED BY GROWTH  
; HORMONE PROMOTER  
; NUMBER OF SEQUENCES: 18  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/528,852  
; FILING DATE: 18-MAY-1990  
; SEQ ID NO: 3:  
; LENGTH: 124  
5223610-3

Query Match 84.6%; Score 33; DB 6; Length 124;  
Best Local Similarity 85.7%; Pred. No. 11;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
||| |||  
Db 72 EVPSSQH 78

RESULT 21  
US-08-255-471-9  
; Sequence 9, Application US/08255471  
; Patent No. 5721096  
; GENERAL INFORMATION:  
; APPLICANT: Karathanasis, Sotirios K.  
; APPLICANT: Ladias, John A.  
; APPLICANT: Rottman, Jeffrey N.  
; APPLICANT: Widom, Russell L.  
; TITLE OF INVENTION: Control of the Apolipoprotein AI Gene  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/255,471  
; FILING DATE: 08-JUN-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/784,472  
; FILING DATE: 24-OCT-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carroll, Alice O.  
; REGISTRATION NUMBER: 33,542  
; REFERENCE/DOCKET NUMBER: CMCC-234A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 861-6240  
; TELEFAX: (617) 861-9540  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 414 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-255-471-9

Query Match 79.5%; Score 31; DB 1; Length 414;  
Best Local Similarity 100.0%; Pred. No. 1e-02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQ 6  
Db 14 EVPGSQ 19  
RESULT 22  
5194375-6  
; Patent No. 5194375  
; APPLICANT: PARK, LINDA S.; GOODWIN, RAYMOND G.  
; TITLE OF INVENTION: DNA ENCODING INTERLEUKIN-7 RECEPTORS AND  
; METHODS OF USE  
; NUMBER OF SEQUENCES: 9  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/493,588  
; FILING DATE: 21-MAY-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 421,201  
; FILING DATE: 13-OCT-1989  
; APPLICATION NUMBER: 366,910  
; FILING DATE: 15-JUN-1989  
; SEQ ID NO: 6:  
; LENGTH: 459  
5194375-6

Query Match 79.5%; Score 31; DB 6; Length 459;  
Best Local Similarity 85.7%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 EVPGSQ 7  
Db 47 EVDGSQ 53

RESULT 23  
5194375-4  
; Patent No. 5194375  
; APPLICANT: PARK, LINDA S.; GOODWIN, RAYMOND G.  
; TITLE OF INVENTION: DNA ENCODING INTERLEUKIN-7 RECEPTORS AND  
; METHODS OF USE  
; NUMBER OF SEQUENCES: 9  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/493,588  
; FILING DATE: 21-MAY-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 421,201  
; FILING DATE: 13-OCT-1989  
; APPLICATION NUMBER: 366,910  
; FILING DATE: 15-JUN-1989  
; SEQ ID NO: 4:  
; LENGTH: 262  
5194375-4

Query Match 76.9%; Score 30; DB 6; Length 262;  
Best Local Similarity 85.7%; Pred. No. 98;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 EVPGSQ 7  
Db 47 EVNGSQ 53

RESULT 24  
US-08-781-802-8  
; Sequence 8, Application US/08781802  
; Patent No. 5969121  
; GENERAL INFORMATION:  
; APPLICANT: ALLEN, Larry  
; APPLICANT: AIKENS, John  
; APPLICANT: FONSTEIN, Michael  
; APPLICANT: VONSTEIN, Veronika  
; APPLICANT: DEMIRJIAN, David

; APPLICANT: CASADABAN, Malcol  
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
; STREET: 300 S. Wacker Drive 32nd Floor  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,802  
; FILING DATE: 10-JAN-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/694,078  
; FILING DATE: 07-AUG-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/019,580  
; FILING DATE: 12-JUN-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/009,704  
; FILING DATE: 11-JAN-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/001,995  
; FILING DATE: 01-AUG-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Chao, Mark  
; REGISTRATION NUMBER: 37,293  
; REFERENCE/DOCKET NUMBER: 95,963-E  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-913-0001  
; TELEFAX: 312-913-0002  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 329 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-781-802-8

Query Match 76.9%; Score 30; DB 2; Length 329;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PGSQ 7  
Db 289 PGSQ 293

RESULT 25  
US-08-694-078-8  
; Sequence 8, Application US/08694078  
; Patent No. 6218163  
; GENERAL INFORMATION:  
; APPLICANT: ALLEN, Larry  
; APPLICANT: AIKENS, John  
; APPLICANT: FONSTEIN, Michael  
; APPLICANT: VONSTEIN, Veronika  
; APPLICANT: DEMIRJIAN, David  
; APPLICANT: CASADABAN, Malcol  
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff, Ltd.  
; STREET: 300 S. Wacker Drive 7th Floor

CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/694,078  
FILING DATE: 07-AUG-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/019,580  
FILING DATE: 12-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/009,704  
FILING DATE: 10-JAN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/001,995  
FILING DATE: 07-AUG-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Chao, Mark  
REGISTRATION NUMBER: 37,293  
REFERENCE/DOCKET NUMBER: 95,963-C  
TELEPHONE: 312-913-0001  
TELEFAX: 312-913-0002  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 329 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-694-078-8

Query Match 76.98; Score 30; DB 4; Length 329;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 3 PGSQH 7  
Db 289 PGSQH 293  
|||||

Search completed: July 16, 2001, 16:36:18  
Job time: 207 sec

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 16:33:31 ; Search time 34.9 seconds  
(without alignments)  
15.279 Million cell updates/sec

Title: US-09-786-648-2

Perfect score: 39

Sequence: 1 EVPGSQH 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 68:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	100.0	124	1	XVVCB
2	39	100.0	124	1	QLECB
3	33	84.6	272	2	T14755
4	33	84.6	367	2	T01751
5	32	82.1	455	1	A63753
6	32	82.1	534	2	S44886
7	31	79.5	137	2	B27586
8	31	79.5	242	2	T34767
9	31	79.5	260	2	C96827
10	31	79.5	273	2	B48820
11	31	79.5	274	2	G84353
12	31	79.5	353	2	H75446
13	31	79.5	374	2	D81715
14	31	79.5	414	2	A37133
15	31	79.5	414	2	I48975
16	31	79.5	432	2	T46725
17	31	79.5	432	2	T51020
18	31	79.5	459	2	D34791
19	31	79.5	489	2	T26069
20	31	79.5	1028	2	A96719
21	31	79.5	1258	2	T29041
22	31	79.5	1259	2	A43425
23	30	76.9	148	2	B72782
24	30	76.9	153	2	T31701
25	30	76.9	182	2	D83638
26	30	76.9	262	2	C34791
27	30	76.9	270	2	D83072
28	30	76.9	298	2	B34791
29	30	76.9	304	2	F84169

30 30 76.9 336 2 A72247  
31 30 76.9 373 2 T47115  
32 30 76.9 375 2 T35015  
33 30 76.9 376 1 S17246  
34 30 76.9 412 2 T47142  
35 30 76.9 436 2 B70321  
36 30 76.9 440 2 T44138  
37 30 76.9 459 2 A34791  
38 30 76.9 500 2 J40422  
39 30 76.9 573 2 JC4335  
40 30 76.9 616 2 I38155  
41 30 76.9 622 2 T22716  
42 30 76.9 623 2 T47874  
43 30 76.9 679 2 T05041  
44 30 76.9 775 2 S41962  
45 30 76.9 805 2 A46266

## ALIGNMENTS

RESULT 1

XVVCB

cholera enterotoxin, B chain precursor VC1456 [validated] - Vibrio cholerae (strain N  
N; Alternate names: enterotoxin beta chain  
C; Species: Vibrio cholerae  
C; Date: 24-Apr-1984 #sequence\_revision 01-Sep-2000 #text\_change 02-Feb-2001  
C; Accession: S14624; S39238; S39241; H82196; JC1078; S17666; PC1010; A05130; A01819;  
R; Dams, E.; de Wolf, M.; Dierick, W.  
submitted to the EMBL Data Library, March 1991

A; Description: Correction of the cholera toxin nucleotide sequence of the Vibrio chol  
A; Reference number: S14623  
A; Accession: S14624  
A; Molecule type: DNA

A; Residues: 1-124 &lt;DAM&gt;

A; Cross-references: EMBL:X58786; NID:g48420; PIDN:CAA41593.1; PID:g48422

A; Experimental source: strain 2125

R; Lebens, W.; Holmgren, J.

submitted to the EMBL Data Library, November 1993

A; Description: Structure and arrangement of the Cholera toxin genes in vibrio Cholera

A; Reference number: S39238

A; Accession: S39238

A; Molecule type: DNA

A; Residues: 1-124 &lt;UEB&gt;

A; Cross-references: EMBL:X76390; NID:g433856; PIDN:CAA53973.1; PID:g433857  
A; Accession: S39241

A; Molecule type: DNA

A; Residues: 1-124 &lt;LEW&gt;

A; Cross-references: EMBL:X76391; NID:g433859; PIDN:CAA53976.1; PID:g433861  
R; Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.  
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A; Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A; Reference number: A82035; MUID:20406833

A; Accession: H82196

A; Status: preliminary

A; Molecule type: DNA

A; Residues: 1-124 &lt;HEI&gt;

A; Cross-references: GB:AE004224; GB:AE003852; NID:g9655952; PIDN:AAF94613.1; GSPDB:GM

A; Experimental source: serogroup O1; strain N16961; biotype El Tor

R; Shi, C.H.; Cao, C.; Zhang, J.S.; Ma, Q.J.

Chinese Biochem. J. 9, 395-399, 1993

A; Title: Nucleotide sequence analysis of the gene encoding the classical biotype V. cl.

A; Reference number: JC1078

A; Accession: JC1078

A; Molecule type: DNA

A; Residues: 1-20, 'Q', '22-31, 'Q', '33-38, 'H', '40-49, 'G', '51-67, 'T', '69-124 &lt;SHI&gt;

A; Experimental source: classical biotype strain 569B

R; Dams, E.; de Wolf, M.; Dierick, W.

Biochim. Biophys. Acta 1090, 139-141, 1991

A; Title: Nucleotide sequence analysis of the CT operon of the Vibrio cholerae classic

A; Reference number: S17665; MUID:91355224

DNA-directed RNA p  
probable 4-carboxy  
probable 3-oxodip  
chorismate synthas  
hypothetical prote  
flavocytochrome C  
hypothetical prote  
interleukin-7 rece  
4-aminobutyrate tr  
anti-mullerian hor  
DNA-binding regula  
hypothetical prote  
scarce-like pro  
hypothetical prote  
thiamin repressibl  
aryl hydrocarbon r

A:Accession: S17666  
A:Molecule type: DNA  
A:Residues: 1-38,'H',40-67,'T',69-124 <DA>  
A:Cross-references: EMBL:X58785; NID:g48888; PIDN:CAA41591.1; PID:g48890  
R:Ma, Q.J.; Liu, C.X.; Xiong, L.S.; Yu, X.Q.  
Sci. Sin. B Chem. Biol. Agric. Med. Earth Sci. 34, 274-280, 1991  
A:Title: B subunit of cholera toxin produced in *Escherichia coli*.  
A:Reference number: PC1010  
A:Accession: PC1010  
A:Molecule type: protein  
A:Residues: 22-38,'H',40-41 <MA>  
R:Mekalanos, J.J.; Swartz, D.J.; Pearson, G.D.N.; Harford, N.; Groyne, F.; de Wilde, M.  
Nature 306, 551-557, 1983  
A:Reference number: A93320; MUID:84068199  
A:Accession: A05130  
A:Molecule type: DNA  
A:Residues: 1-32,'S',34-74,'S',76-124 <ME>  
A:Cross-references: GB:X00171; NID:g48347; PIDN:CAA24996.1; PID:g758351  
R:Kurosky, A.; Markel, D.E.; Peterson, J.W.  
J. Biol. Chem. 252, 7257-7264, 1977  
A:Title: Covalent structure of the beta chain of cholera enterotoxin.  
A:Reference number: A01819; MUID:78005537  
A:Accession: A01819  
A:Molecule type: protein  
A:Residues: 22-38,'H',40-42,'N',44-67,'T',69-90,'N',92-124 <KUR>  
R:Lai, C.Y.  
J. Biol. Chem. 252, 7249-7256, 1977  
A:Title: Determination of the primary structure of cholera toxin B subunit.  
A:Reference number: A38033; MUID:78005536  
A:Accession: A38033  
A:Molecule type: protein  
A:Residues: 22-38,'H',40-42,'N',44-67,'T',69,'E',71-90,'N',92-124 <LAI>  
A:Note: the difference at residue 70 may be due to deamidation during preparation  
R:Nakashima, Y.; Napiorkowski, P.; Schafer, D.E.; Konigsberg, W.H.  
FEBS Lett. 68, 275-278, 1976  
A:Title: Primary structure of the B subunit of cholera enterotoxin.  
A:Reference number: A38034; MUID:77026365  
A:Accession: A38034  
A:Molecule type: protein  
A:Residues: 22-38,'H',40-67,'T',69,'E',71,'QS',74-75,'VE',78-86,'Q',88-99,'Q',101-103,'Q'  
R:Takao, T.; Watanabe, H.; Shimonishi, Y.  
Eur. J. Biochem. 146, 503-508, 1985  
A:Title: Facile identification of protein sequences by mass spectrometry.  
A:Reference number: A21910; MUID:85126976  
A:Accession: A21910  
A:Molecule type: protein  
A:Residues: 22-38,'H',40-42,'N',44-67,'T',69-90,'N',92-124 <TAK>  
A:Experimental source: biotype Inaba 569B  
A:Note: Asn-65 was partially deaminated to Asp  
C:Comment: The authors translated the codon TCA for residue 33 as Tyr.  
C:Genetics:  
A:Gene: VC1456  
A:Map position: 1  
C:Complex: the cholera enterotoxin molecule contains three kinds of chains; an alpha and  
cysteine noncovalently with the subunit B, an aggregate of five beta chains  
C:Function:  
C:Superfamily: involved in binding of the toxin to cell membranes  
C:Superfamily: cholera enterotoxin beta chain  
C:Keywords: enterotoxin; toxin  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-124/Product: cholera enterotoxin chain B #status predicted <MAT>  
F:30-107/Disulfide bonds: #status experimental

Query Match 100.0%; Score 39; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 0.38;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EVPGSQH 7

|||||||

Db 72 EVPGSQH 78

RESULT 2  
OLECB  
Heat-labile enterotoxin chain B precursor - *Escherichia coli*  
C:Species: *Escherichia coli*  
C>Date: 29-Jun-1981 sequence\_revision 29-Jun-1991 #text\_change 18-Jun-1999  
C:Accession: A01820; B26946; I41194; I41287; I67644; A61475  
R:Dallas, W.S.; Falkow, S.  
Nature 288, 499-501, 1980  
A:Title: Amino acid sequence homology between cholera toxin and *Escherichia coli* heat-labile enterotoxin  
A:Reference number: A01820; MUID:81074965  
A:Accession: A01820  
A:Molecule type: mRNA  
A:Residues: 1-124 <DAL>  
R:Yamamoto, T.; Gojohori, T.; Yokota, T.  
J. Bacteriol. 169, 1352-1357, 1987  
A:Title: Evolutionary origin of pathogenic determinants in enterotoxigenic *Escherichia coli*  
A:Reference number: A26946; MUID:87137303  
A:Accession: B26946  
A:Molecule type: DNA  
A:Residues: 1-27,'E',29-63,'K',65-124 <YAM>  
A:Cross-references: EMBL:M15363; NID:g148335; PIDN:AAA24792.1; PID:g148336  
R:Leong, J.; Vinal, A.C.; Dallas, W.S.  
Infect. Immun. 48, 73-77, 1985  
A:Title: Nucleotide sequence comparison between heat-labile toxin B-subunit cistrons  
A:Reference number: I41194; MUID:85156481  
A:Accession: I41194  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-5,'F',7-17,'C',19-24,'S',26-27,'E',29-33,'H',35-63,'K',65-66,'A',68-124  
A:Cross-references: GB:M17874; NID:g145830; PIDN:AAA98064.1; PID:g145831  
R:Ibrahim, I.; Gentz, R.  
J. Biol. Chem. 262, 10189-10194, 1987  
A:Title: A functional interaction between the signal peptide and the translation apparatus.  
A:Reference number: I41287; MUID:87280041  
A:Accession: I41287  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-22 <RE2>  
A:Cross-references: GB:M17101; NID:g146375; PIDN:AAA23973.1; PID:g146376  
R:Pinoue, T.; Tsuji, T.; Koto, M.; Imamura, S.; Miyama, A.  
FEMS Microbiol. Lett. 108, 157-161, 1993  
A:Title: Amino acid sequence of heat-labile enterotoxin from chicken enterotoxigen  
A:Reference number: I53542; MUID:93252225  
A:Accession: I57644  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-17,'C',19,'Y',21-24,'S',26-27,'E',29-63,'K',65-66,'A',68-122,'E',124 <TSU>  
A:Cross-references: GB:S60731; NID:g408994; PIDN:AAC60441.1; PID:g408996  
R:Tsuji, T.; Lida, T.; Honda, T.; Miwatani, T.; Nagehama, M.; Sakurai, J.; Wada, K.  
Microb. Pathog. 2, 381-390, 1987  
A:Title: A unique amino acid sequence of the B subunit of a heat-labile enterotoxin  
A:Reference number: A61475; MUID:89180953  
A:Accession: A61475  
A:Molecule type: protein  
A:Residues: 22-24,'S',26-27,'E',29-63,'K',65-66,'A',68-95,'A',97-122,'E',124 <TSU>  
A:Experimental source: strain 240-3  
C:Complex: the heat-labile enterotoxin molecule contains one A chain and five or six  
C:Function:  
C:Superfamily: cholera enterotoxin beta chain  
C:Keywords: enterotoxin  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-124/Product: heat-labile enterotoxin chain B #status predicted <MAT>  
F:30-107/Disulfide bonds: #status predicted

Query Match 100.0%; Score 39; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 0.38;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EVPGSQH 7



Db 72 EYPGSOH 78  
|||||

# RESULT 3

T14755 hypothetical protein DKFZp564A0122.1 - human

C:Species: Homo sapiens (man)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T14755  
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, August 1999  
A:Reference number: Z18181  
A:Accession: T14755  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-272 <WAM>  
A:Cross-references: EMBL:AL110209  
A:Experimental source: fetal brain; clone DKFZp564A0122  
C:Genetics:  
A:Note: DKFZp564A0122.1

Query Match 84.6%; Score 33; DB 2; Length 272;  
Best Local Similarity 71.4%; Pred. No. 17;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
I:||||:  
Db 246 ELPGSEH 252

# RESULT 4

T01751 gibberellin 20-oxidase - common tobacco  
N:Alternate names: Ntcl6 protein  
C:Species: Nicotiana tabacum (common tobacco)  
C:Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 20-Jun-2000  
C:Accession: T01751  
R:Tanaka-Ueguchi, M.; Itoh, H.; Oyama, N.; Koshioka, M.; Matsuoka, M.  
submitted to the EMBL Data Library, July 1998  
A:Description: Over-expression of a tobacco homeobox gene, NTH15, decreases the expression of gibberellin 20-oxidase.  
A:Reference number: Z14418  
A:Accession: T01751  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-367 <TAN>  
A:Cross-references: EMBL:AB016084  
C:Genetics:  
A:Gene: Ntcl6  
C:Superfamily: 1-aminocyclopropane-1-carboxylate oxidase

Query Match 84.6%; Score 33; DB 2; Length 367;  
Best Local Similarity 85.7%; Pred. No. 23;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
I:||||:  
Db 149 EYPSSOH 155

# RESULT 5

A69753 glucarate dehydratase (EC 4.2.1.40) - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
C:Accession: A69753  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertoni, C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chodura, A.; Ehrlich, S.D.; Emmeron, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fritz, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Geller

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M. Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl A:Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Se akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida A:Authors: Yoshikawa, H.F.; zumstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis  
A:Reference number: A69580; MUID:98044033  
A:Accession: A69753  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-455 <KUN>  
A:Cross-references: GB:Z99105; GB:AL009126; NID:g2632457; PIDN:CAB12043.1; PID:g26325  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: ycbF  
C:Superfamily: glucarate dehydratase  
C:Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 82.1%; Score 32; DB 1; Length 455;  
Best Local Similarity 71.4%; Pred. No. 48;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
I:||||:  
Db 64 EYPGGEH 70

# RESULT 6

S44886 112.1 protein - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 14-Sep-1994 #sequence\_revision 12-May-1995 #text\_change 23-Mar-2001  
C:Accession: S44886  
R:Du, Z.  
submitted to the EMBL Data Library, May 1993  
A:Description: Sequence of the C. elegans cosmid ZK112.  
A:Reference number: S44616  
A:Accession: S44886  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-534 <DUZ>  
A:Cross-references: EMBL:L14324; NID:g289740; PID:g289741  
C:Genetics:  
A:Introns: 25/3; 65/2; 196/2; 249/1; 275/1; 385/2; 415/2  
C:Superfamily: Caenorhabditis elegans ZK688.6 protein

Query Match 82.1%; Score 32; DB 2; Length 534;  
Best Local Similarity 71.4%; Pred. No. 57;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
I:||||:  
Db 442 EIPGSAH 448

# RESULT 7

B27586 hypothetical protein - Mycobacterium leprae  
C:Species: Mycobacterium leprae  
C:Date: 08-Mar-1989 #sequence\_revision 08-Mar-1989 #text\_change 26-May-2000  
C:Accession: B27586  
R:Booth, K.J.; Harris, D.P.; Love, J.M.; Watson, J.D.  
J. Immunol. 140, 597-601, 1988  
A:Title: Antigenic proteins of Mycobacterium leprae. Complete sequence of the gene fo  
A:Reference number: A92821; MUID:88088878  
A:Accession: B27586  
A:Molecule type: DNA

A:Residues: 1-137 <BO>  
A:Cross-references: GB:M19058; NID:gl149919; PIDN:AAA88230.1; PID:gl1196505  
C:Superfamily: Mycobacterium leprae hypothetical 15.2K protein

Query Match 79.5%; Score 31; DB 2; Length 137;  
Best Local Similarity 83.3%; Pred. No. 22;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 VPGSQH 7  
|||:|  
DB 127 VPGNQH 132

RESULT 8  
T34767  
hypothetical protein SC2A11.21c SC2A11.21c - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
C:Accession: T34767  
R:Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, August 1998  
A:Reference number: Z21536  
A:Accession: T34767  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-242 <MUR>  
A:Cross-references: EMBL:AL031184; PIDN:CAA20190.1; GSPDB:GN00070; SCOEDB:SC2A11.21c  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC2A11.21c

Query Match 79.5%; Score 31; DB 2; Length 242;  
Best Local Similarity 71.4%; Pred. No. 40;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 EVPGSQH 7  
||||:|  
DB 205 EVPGIDH 211

RESULT 9  
C96827  
protein F20B17.2 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: C96827  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: C96827  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-260 <STO>  
A:Cross-references: GB:AE005173; NID:g7715588; PIDN:AAF68106.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F20B17.2  
A:Map position: 1

Query Match 79.5%; Score 31; DB 2; Length 260;  
Best Local Similarity 71.4%; Pred. No. 43;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
:|||||  
DB 66 KVPCKQH 72

RESULT 10  
B48820  
homeobox protein (clone NVHBox-4) - eastern newt  
C:Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)  
C:Date: 01-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 24-Sep-1999  
C:Accession: B48820  
R:Beauchemin, M.; Savard, P.  
Dev. Biol. 154, 55-65, 1992  
A:Title: Two distal-less related homeobox-containing genes expressed in regeneration  
A:Reference number: A48820; MUID:93050784  
A:Accession: B48820  
A:Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-273 <BEA>  
A:Cross-references: GB:X63531; GB:S47223; NID:g432377; PIDN:CAA45094.1; PID:g432378  
A:Note: sequence extracted from NCBI backbone (NCBIN:117052, NCBIP:117053)  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:126-182/Domain: homeobox homology <HOX>

Query Match 79.5%; Score 31; DB 2; Length 273;  
Best Local Similarity 71.4%; Pred. No. 45;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
||||:|  
DB 188 EVPGMEH 194

RESULT 11  
G84353  
hypothetical protein Vng2034h [imported]. - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: G84353  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky,  
Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483  
A:Accession: G84353  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-274 <STO>  
A:Cross-references: GB:AE004437; NID:gl0581460; PIDN:AAG20195.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: VNG2034H

Query Match 79.5%; Score 31; DB 2; Length 274;  
Best Local Similarity 71.4%; Pred. No. 46;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
|||||  
DB 117 EAPGDQH 123

RESULT 12  
H75446  
(S)-2-hydroxy-acid oxidase - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 28-Jul-2000  
C:Accession: H75446  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M.  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896

A:Accession: H75446

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-353 <WHI>

A:Cross-references: GB:AE001954; GB:AE000513; NID:g6458751; PIDN:AAF10604.1; PID:g645875

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR1031

A:Map position: 1

C:Superfamily: (S)-2-hydroxy-acid oxidase; (S)-2-hydroxy-acid oxidase homology

F:3-297/Domain: (S)-2-hydroxy-acid oxidase homology <2HY>

Query Match 79.5%; Score 31; DB 2; Length 353;

Best Local Similarity 83.3%; Pred. NO. 60;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VPGSQH 7

Db 189 VPGSEH 194

|||||

RESULT 13

DB1715

conserved hypothetical protein TC0328 [imported] - Chlamydia muridarum (strain Nigg)

C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn

C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000

C:Accession: DB1715

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,

, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.

A:Reference number: AB1500; MUID:20150255

A:Accession: DB1715

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-374 <TET>

A:Cross-references: GB:AE002300; GB:AE002160; NID:g7190365; PIDN:AAF39192.1; PID:g719037

A:Experimental source: strain Nigg (MoPn)

C:Genetics:

A:Gene: TC0328

Query Match 79.5%; Score 31; DB 2; Length 374;

Best Local Similarity 83.3%; Pred. NO. 63;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VPGSQH 7

Db 327 LPGSQH 332

|||||

RESULT 14

A37133

apolipoprotein A-I regulatory protein 1 - human

N:Alternate names: ARP-1; ovalbumin upstream promoter transcription factor II

C:Species: Homo sapiens (man)

C>Date: 08-Mar-1991 #sequence\_revision 08-Mar-1991 #text\_change 20-Sep-1999

C:Accession: A37133; I54072

R:Ladakis, J.A.A.; Karathanasis, S.K.

Science 251, 561-565, 1991

A:Title: Regulation of the apolipoprotein AI gene by ARP-1, a novel member of the steroid

A:Reference number: A37133; MUID:91118002

A:Accession: A37133

A:Molecule type: mRNA

A:Residues: 1-414 <LAD>

A:Cross-references: GB:M64497; NID:g179023; PIDN:AAAB6429.1; PID:g179024

R:Wang, L.H.; Ing, N.H.; Tsai, S.Y.; O'Malley, B.W.; Tsai, M.J.

chorismate synthase (EC 4.6.1.4) / flavin reductase, NADPH-dependent [validated] - Ne

Gene Expr. 1, 207-216, 1991

A:Title: The COUP-TFs compose a family of functionally related transcription factors.

A:Reference number: I54072; MUID:92314709

A:Accession: I54072

A>Status: translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-351 <RES>

A:Cross-references: GB:M62760; NID:g180930; PIDN:AAA21479.1; PID:g538262

C:Genetics:

A:Gene: GDB:TFCOUP2; ARP1

A:Cross-references: GDB:511233; OMIM:107773

A:Map position: 15q26.1-15q26.2

C:Superfamily: unassigned erba-related proteins; erba transforming protein homology

C:Keywords: DNA binding; lipid binding; lipoprotein; transcription factor; zinc finger

F:77-323/Domain: erba transforming protein homology <ERBA>

F:79-99/Region: zinc finger

F:115-139/Region: zinc finger

Query Match 79.5%; Score 31; DB 2; Length 414;

Best Local Similarity 100.0%; Pred. NO. 71;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQ 6

Db 14 EVPGSQ 19

|||||

RESULT 15

I48975

apolipoprotein A-I regulatory protein 1 - mouse

N:Alternate names: ARP-1 protein; COUP-TFII

C:Species: Mus musculus (house mouse)

C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 20-Sep-1999

C:Accession: I48975; I48732; S44284

R:Qiu, Y.; Cooney, A.J.; Kuratani, S.; DeMayo, F.J.; Tsai, S.Y.; Tsai, M.J.

Proc. Natl. Acad. Sci. U.S.A. 91, 4451-4455, 1994

A:Title: Spatiotemporal expression patterns of chicken ovalbumin upstream promoter-tr

the dlencephalon.

A:Reference number: I48973; MUID:94240155

A:Accession: I48975

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-414 <RES>

A:Cross-references: EMBL:U07635; NID:g466469; PIDN:AAA13854.1; PID:g466470

R:Jonk, L.J.; de Jonge, M.E.; Pals, C.E.; Wissink, S.; Vervaart, J.M.; Schoorlemmer,

Mech. Dev. 47, 81-97, 1994

A:Title: Cloning and expression during development of three murine members of the COU

A:Reference number: I48305; MUID:95034311

A:Accession: I48732

A>Status: translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-35, 'p', 37-414 <RE2>

A:Cross-references: EMBL:X76653; NID:g482927; PIDN:CAA54096.1; PID:g482928

C:Genetics:

A:Gene: ARP-1

C:Superfamily: unassigned erba-related proteins; erba transforming protein homology

F:77-323/Domain: erba transforming protein homology <ERBA>

Query Match 79.5%; Score 31; DB 2; Length 414;

Best Local Similarity 100.0%; Pred. NO. 71;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQ 6

Db 14 EVPGSQ 19

|||||

RESULT 16

T46725

chorismate synthase (EC 4.6.1.4) / flavin reductase, NADPH-dependent [validated] - Ne

N:Contains: chorismate synthase (EC 4.6.1.4): flavin reductase, NADPH-dependent  
C:Species: Neurospora crassa  
C:Date: 17-Mar-2000 #sequence\_revision 17-Mar-2000 #text\_change 02-Sep-2000  
C:Accession: T46725  
R:Henstrand, J.M.; Amrhein, N.; Schmid, J.  
J. Biol. Chem. 270, 20447-20452, 1995  
A:Title: Cloning and Characterization of a Heterologously Expressed Bifunctional Chorismate Synthase  
A:Reference number: Z06450; MUID:95386486  
A:Accession: T46725  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-432 <H>  
A:Cross-references: EMBL:U25818; NID:9976374; PIDN:AAC49056.1; PID:9976375  
C:Function:  
A:Description: EC 4.6.1.4 [validated; MUID:95386486]  
A:Note: N. crassa chorismate synthase (CS) is a bifunctional enzyme, possessing also an FT activity  
C:Function: <PRE>  
A:Note: CS and FT activity are encoded by a single domain  
A:Description: flavin reductase [validated; MUID:95386486]  
A:Note: N. crassa chorismate synthase (CS) is a bifunctional enzyme, possessing also an FT activity  
C:Function: <PRE>  
A:Note: CS and FT activity are encoded by a single domain  
C:Superfamily: chorismate synthase  
C:Keywords: aromatic amino acid biosynthesis; phosphorus-oxygen lyase; shikimate pathway

Query Match 79.5%; Score 31; DB 2; Length 432;  
Best Local Similarity 85.7%; Pred. No. 74;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVPGSQH 7  
|||||  
Db 274 EVPGSIH 280

RESULT 17  
T51020  
chorismate synthase/flavin reductase, NADPH-dependent [imported] - Neurospora crassa  
N:Alternate names: protein B7F21.10  
C:Species: Neurospora crassa  
C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 02-Sep-2000  
C:Accession: T51020  
R:Schulte, U.; Align, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, July 2000  
A:Reference number: Z25286  
A:Accession: T51020  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-432 <SCH>  
A:Cross-references: EMBL:AL389901; GSPDB:GN00116; NCSP:B7F21.10  
A:Experimental source: BAC clone B7F21; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B7F21.10  
A:Map position: 6  
A:Introns: 12/2; 92/1  
C:Superfamily: chorismate synthase

Query Match 79.5%; Score 31; DB 2; Length 432;  
Best Local Similarity 85.7%; Pred. No. 74;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVPGSQH 7  
|||||  
Db 274 EVPGSIH 280

RESULT 18  
D34791  
interleukin-7 receptor precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 23-Jul-1999  
C:Accession: D34791; C40256  
R:Goodwin, R.G.; Friend, D.; Ziegler, S.F.; Jerzy, R.; Falk, B.A.; Gimpel, S.; Cosman, D

Cell 60, 941-951, 1990  
A:Title: Cloning of the human and murine interleukin-7 receptors: demonstration of a reference sequence  
A:Reference number: A34791; MUID:90199875  
A:Accession: D34791  
A:Molecule type: mRNA  
A:Residues: 1-459 <GOO>  
A:Cross-references: GB:M29697; NID:9198377; PIDN:AAA39304.1; PID:9309411  
R:Pleiman, C.M.; Gimpel, S.D.; Park, L.S.; Harada, H.; Taniguchi, T.; Ziegler, S.F.  
Mol. Cell. Biol. 11, 3052-3059, 1991  
A:Title: Organization of the murine and human interleukin-7 receptor genes: two mRNAs  
A:Reference number: A40256; MUID:91246172  
A:Accession: C40256  
A:Molecule type: DNA  
A:Residues: 231-239; 264-272 <PLE>  
C:Superfamily: interleukin-7 receptor; fibronectin type III repeat homology  
C:Keywords: cytokine receptor; phosphoprotein; transmembrane protein  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-459/Product: interleukin-7 receptor #status predicted <MAT>  
F:21-239/Domain: extracellular #status predicted <EXT>  
F:240-264/Domain: transmembrane #status predicted <TMM>

Query Match 79.5%; Score 31; DB 2; Length 459;  
Best Local Similarity 85.7%; Pred. No. 79;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVPGSQH 7  
|||||  
Db 47 EVDGSIH 53

RESULT 19  
T26069  
hypothetical protein W02A11.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 02-Sep-2000  
C:Accession: T26069  
R:White, S.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z20147  
A:Accession: T26069  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-489 <WIL>  
A:Cross-references: EMBL:Z82062; PIDN:CAB04890.1; GSPDB:GN00019; CESP:W02A11.3  
A:Experimental source: clone W02A11  
C:Genetics:  
A:Gene: CESP:W02A11.3  
A:Map position: 1  
A:Introns: 58/3; 129/3; 361/3; 444/3  
C:Superfamily: RING finger homology  
F:429-479/Domain: RING finger homology <RRN>

Query Match 79.5%; Score 31; DB 2; Length 489;  
Best Local Similarity 83.3%; Pred. No. 85;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VPGSQH 7  
|||||  
Db 218 VPGSEH 223

RESULT 20  
A96719  
hypothetical protein T6C23.9 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: A96719  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Mailli, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: A96719  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1028 <STO>  
A:Cross-references: GB:AE005173; NID:g6665545; PIDN:AAF22914.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: T6C23.9  
A:Map position: 1

Query Match 79.5%; Score 31; DB 2; Length 1028;  
Best Local Similarity 83.3%; Pred. No. 1.9e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VPGSQH 7  
|||||  
Db 636 VPGSEH 641

RESULT 21  
T29041  
hypothetical protein B0228.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
C:Accession: T29041  
R:Leinbach, D.  
submitted to the EMBL Data Library, March 1995  
A:Description: The sequence of C. elegans cosmid B0228.  
A:Reference number: Z18324  
A:Accession: T29041  
A>Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-1258 <LEI>  
A:Cross-references: EMBL:U23168; PIDN:AAC38813.1; CESP:B0228.4  
A:Experimental source: strain Bristol N2  
C:Genetics:  
A:Gene: CESP:B0228.4  
A:Introns: 149/3; 299/2; 393/3; 435/1; 586/1; 615/3; 655/3; 678/3; 701/3; 1213/3  
C:Superfamily: Caenorhabditis elegans hypothetical protein B0228.4

Query Match 79.5%; Score 31; DB 2; Length 1258;  
Best Local Similarity 85.7%; Pred. No. 2.3e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
|||||  
Db 195 EVPGSTH 201

RESULT 22  
A43425  
Bravo/Nr-CAM cell adhesion molecule L1 homolog - chicken (fragment)  
C:Species: Gallus gallus (chicken)  
C>Date: 27-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 16-Jul-1999  
C:Accession: A43425  
R:Kayem, J.F.; Roman, J.M.; de la Rosa, E.J.; Schwarz, U.; Dreyer, W.J.  
J. Cell Biol. 118, 1259-1270, 1992  
A:Title: Bravo/Nr-CAM is closely related to the cell adhesion molecules L1 and Ng-CAM and  
A:Reference number: A43425; MUID:92381110  
A:Accession: A43425  
A>Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid; protein  
A:Residues: 1-1259 <KAY>  
A:Experimental source: cerebellum  
A>Note: sequence extracted from NCBI backbone (NCBIP:112026)

C:Superfamily: neural cell adhesion molecule L1; fibronectin type III repeat homology  
F:237-294/Domain: immunoglobulin homology <IMM>

Query Match 79.5%; Score 31; DB 2; Length 1259;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQ 6  
|||||  
Db 651 EVPGSQ 656

RESULT 23  
B72782  
hypothetical protein APE0243 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jun-2000  
C:Accession: B72782  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero  
A:Reference number: A72450; MUID:99310339  
A:Accession: B72782  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-148 <KAW>  
A:Cross-references: DDBJ:AP000058; NID:g5103388; PIDN:BAA79156.1; PID:d1042932; PID:g  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE0243  
C:Superfamily: Aeropyrum pernix hypothetical protein APE0243

Query Match 76.9%; Score 30; DB 2; Length 148;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PGSQH 7  
|||||  
Db 8 PGSQH 12

RESULT 24  
T31701  
hypothetical protein F29A7.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T31701  
R:Waterston, R.; Le, T.T.  
submitted to the EMBL Data Library, July 1997  
A:Description: The sequence of C. elegans cosmid F29A7.  
A:Reference number: Z21071  
A:Accession: T31701  
A>Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-153 <WAT>  
A:Cross-references: EMBL:AF016416; PIDN:AAB65273.1; GSPDB:GN000020; CESP:F29A7.3  
A:Experimental source: strain Bristol N2; clone F29A7  
C:Genetics:  
A:Gene: CESP:F29A7.3  
A:Map position: 2  
A:Introns: 24/3; 109/1

Query Match 76.9%; Score 30; DB 2; Length 153;  
Best Local Similarity 71.4%; Pred. No. 40;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
|||||  
Db 18 EAPGSSH 24

RESULT 25  
D83638  
conserved hypothetical protein PA0054 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: D83638  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:2043737  
A:Accession: D83638  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-182 <STO>  
A:Cross-references: GB:AE004445; GB:AE004091; NID:g9945872; PIDN:AAG03444.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0054

Query Match 76.9%; Score 30; DB 2; Length 182;  
Best Local Similarity 83.3%; Pred. NO. 48;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VFGSQH 7  
||||:|  
Db 118 VFGSRH 123

Search completed: July 16, 2001, 16:37:00  
Job time: 209 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 16:35:46 ; Search time 57.41 Seconds  
(without alignments)  
16.132 Million cell updates/sec

Title: US-09-786-648-2  
Perfect score: 39  
Sequence: 1 EVPGSQH 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_16:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mhc:  
8: sp\_organelle:  
9: sp\_phage:  
10: sp\_plant:  
11: sp\_rodent:  
12: sp\_unclassified:  
13: sp\_vertebrate:  
14: sp\_virus:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	100.0	103	2 Q9R646	Q9r646 vibrio chol
2	39	100.0	124	2 Q57193	Q57193 vibrio chol
3	39	100.0	124	2 Q56635	Q56635 vibrio chol
4	39	100.0	124	2 Q9R215	Q9r215 vibrio chol
5	35	89.7	565	10 Q22511	Q22511 vitis vinif
6	34	87.2	195	13 Q9W7D3	Q9w7d3 oryzias lat
7	34	87.2	199	13 Q9W7D4	Q9w7d4 oryzias lat
8	34	87.2	201	2 Q9R28	Q9r28 streptomyce
9	33	84.6	192	4 Q9NP06	Q9npq6 homo sapien
10	33	84.6	272	4 Q9UG04	Q9ug04 homo sapien
11	33	84.6	367	10 Q80418	Q80418 nicotiana t
12	33	84.6	412	4 Q9Y2B3	Q9y2b3 homo sapien
13	33	84.6	868	5 Q9VFE2	Q9vfe2 drosophila
14	32	82.1	95	10 Q9FQ18	Q9fq18 amaranthus
15	32	82.1	765	11 Q70284	Q70284 mus musculus
16	31	79.5	95	10 Q9FQ38	Q9fq38 amaranthus
17	31	79.5	122	5 Q9K2V0	Q9k2v0 drosophila
18	31	79.5	222	11 Q9JL95	Q9jl95 mus musculus
19	31	79.5	242	2 Q86582	Q86582 streptomyce

20	31	79.5	243	11 Q9QY63	Q9qy63 mus musculus
21	31	79.5	260	10 Q9MA16	Q9ma16 arabidopsis
22	31	79.5	274	1 Q9HNM5	Q9hum5 halobacteri
23	31	79.5	331	5 Q18391	Q18391 drosophila
24	31	79.5	353	2 Q9RVJ7	Q9rvj7 deinococcus
25	31	79.5	371	2 Q9RH15	Q9rh15 z. m
26	31	79.5	374	2 Q9PKY2	Q9pkv2 chlamydia m
27	31	79.5	382	5 Q9V7M7	Q9v7m7 drosophila
28	31	79.5	432	3 Q9P3J3	Q9p3j3 neurospora
29	31	79.5	459	11 Q9R0C1	Q9rdcl mus musculus
30	31	79.5	460	10 Q9LIP8	Q9lif8 arabidopsis
31	31	79.5	489	5 Q9XOM8	Q9xum8 caenorhabdi
32	31	79.5	510	4 Q9H846	Q9h846 homo sapien
33	31	79.5	1258	5 Q10908	Q10908 caenorhabdi
34	30	76.9	65	10 Q9M4H3	Q9m4h3 vitis vinif
35	30	76.9	101	14 Q9YKP7	Q9ykp7 human immun
36	30	76.9	135	4 Q9P2Q7	Q9p2q7 homo sapien
37	30	76.9	148	1 Q9YFK5	Q9yfk5 aeropyrum p
38	30	76.9	153	5 Q16208	Q16208 caenorhabdi
39	30	76.9	154	2 Q68328	Q68328 pseudomonas
40	30	76.9	176	2 Q9FD33	Q9fd33 pseudomonas
41	30	76.9	204	13 Q9W7D2	Q9w7d2 oryzias lat
42	30	76.9	204	13 Q9W7D1	Q9w7d1 oryzias lat
43	30	76.9	209	13 Q9W7D0	Q9w7d0 oryzias lat
44	30	76.9	213	11 P97762	P97762 mus musculus
45	30	76.9	213	13 Q9PTQ5	Q9ptq5 oryzias lat.

ALIGNMENTS

RESULT 1  
Q9R646 ID AC Q9R646 PRELIMINARY; PRT; 103 AA.  
Q9R646 AC Q9R646;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 13, Last annotation update)  
DE CHOLERA-LIKE ENTEROTOXIN B SUBUNIT.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=95303036; PubMed=7783690;  
RA Nakashima K., Eguchi Y., Nakasone N.;  
RT "Characterization of an enterotoxin produced by Vibrio cholerae  
RT O139.";  
RL Microbiol. Immunol. 39:87-94(1995).  
DR HSP; P01556; IXTC.  
DR InterPro: IPR001835;  
DR Pfam: PF01376; Enterotoxin\_B; 1.  
DR PRINTS: PR00772; ENTEROTOXINB.  
SQ SEQUENCE 103 AA; 11645 MW; 992A05C077F2F70E CRC64;

Query Match 100.0%; Score 39; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.71; 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0;

QY 1 EVPGSQH 7  
Db 51 EVPGSQH 57  
|||||||

RESULT 2  
Q57193 ID Q57193 PRELIMINARY; PRT; 124 AA.  
Q57193 AC Q57193;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE CHOLERA TOXIN B PROTEIN (CTB) PRECURSOR (CTB).

9.786-648

RESULT	3
Q56635	
ID	Q56635 PRELIMINARY; PRT; 124 AA.
AC	O56635;
DT	01-NOV-1996 (TREMBLrel. 01, Created)
DD	01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE	01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE	CHOLERA TOXIN PRECURSOR.
GN	CTXB
OS	Vibrio cholerae.
OC	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX	NCBI_TaxID=566;
LN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=S7;
RR	Yamamoto K., Do V.G., Xu M., Iida T., Miwatani T., Albert M.J., Honda T.;
RRT	"Comparison of cholera toxin genes (ctxAB) of non-O1 vibrio cholera strains 854 (O139-bengal) and S7 (O37) from two outbreaks.";
RRL	Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
DR	EMBL; D30052; BAA06289.1; -
DR	HSSP; P01556; 2CBB.
DR	InterPro; IPR001835; -
DR	Pfam; PF01376; Enterotoxin_B; 1.
DR	PRINTS; PR00772; ENTEROTOXINB.
DR	PRODOM; PD012805; -; 1.
DR	SIGNAL.
FT	1 21 POTENTIAL.

5	022511	PRELIMINARY;	PRT;	565 AA.	
AC	022511;				
DT	01-JAN-1998	(TMBLrel. 05, Created)			
DT	01-JAN-1998	(TMBLrel. 05, Last sequence update)			
DT	01-MAR-2001	(TMBLrel. 16, Last annotation update)			
DE	GLUTATHIONE REDUCTASE (NADPH)	(EC 1.6.4.2) (FRAGMENT).			
GN	GOR.				
OS	Vitis vinifera (Grape).				
OS	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta				
OX	Magnoliophyta; eudicotyledons; core eudicots; Vitaceae; Vitis.				
OX	NCBI_TaxID=29760;				
RP	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CV. THOMPSON SEEDLESS (CLONE 2A); TISSUE=FRUIT; ...				
RA	Cassol T., Adams D.O.;				
RA	Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.				
CC	-!- COFACTOR: FAD (BY SIMILARITY).				
CC	-!- SIMILARITY: TO PYRIDINE NUCLEOTIDE-DISULPHIDE OXIDOREDUCTASES				
CC	CLASS-I				
CC	EMBL; AF019907; AAB70837.1; ..				
DR	HSP; P00390; IALG.				
DR	Mendel; 26381; Vitvi.1190; 26381.				
DR	InterPro; IPR000103; ..				



DR InterPro: IPR001100; -

DR Pfam: PF00070; PYL-redox; 1.

DR PRINTS: PR00368; FADPNR.

DR PRINTS: PR00411; PNDRTASEI.

DR PRINTS: PR00469; PNDRTASEII.

DR PROSITE: PS00076; PYRIDINE-REDOX\_1; 1.

KW FAD; Flavoprotein; Oxidoreductase; Redox-active center.

FT NON\_TER 1

SQ SEQUENCE 565 AA; 60695 MW; B26113AE09A121DE CRC64;

Query Match 89.7%; Score 35; DB 10; Length 565;

Best Local Similarity 71.4%; Pred. No. 28;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7

Db 239 EIPGSEH 245

RESULT 6

Q9W7D3

ID Q9W7D3 PRELIMINARY; PRT; 195 AA.

DT 01-NOV-1999 (TRENBLrel. 12, Created)

DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)

DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)

DE HYPOTHETICAL 21.4 KDA PROTEIN.

OS Oryzias latipes (Medaka fish).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;

OC Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.

OX NCBI\_TaxID=8090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ORANGE-RED;

RA Kanamori A.; identification of genes expressed during early oogenesis

RT "Systematic identification of genes expressed during early oogenesis

RT in medaka.";

RL Mol. Reprod. Dev. 0:0-0(1999).

DR EMBL; AF128818; AAD38915.1; -

KW Hypothetical protein.

SQ SEQUENCE 195 AA; 21446 MW; B90682D695729E88 CRC64;

Query Match

Best Local Similarity 100.0%; Pred. No. 15;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VPGSQH 7

Db 111 VPGSQH 116

RESULT 7

Q9W7D4

ID Q9W7D4 PRELIMINARY; PRT; 199 AA.

DT 01-NOV-1999 (TRENBLrel. 12, Created)

DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)

DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)

DE HYPOTHETICAL 21.7 KDA PROTEIN.

OS Oryzias latipes (Medaka fish).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;

OC Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.

OX NCBI\_TaxID=8090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ORANGE-RED;

RA Kanamori A.; identification of genes expressed during early oogenesis

RT "Systematic identification of genes expressed during early oogenesis

RT in medaka.";

RL Mol. Reprod. Dev. 0:0-0(1999).

DR EMBL; AF128817; AAD38914.1; -

KW Hypothetical protein.

SQ SEQUENCE 199 AA; 21726 MW; E9827C05451B15CD CRC64;

QY 2 VPGSQH 7

Db 115 VPGSQH 120

RESULT 8

Q9RJZ8

ID Q9RJZ8 PRELIMINARY; PRT; 201 AA.

DT 01-MAY-2000 (TRENBLrel. 13, Created)

DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)

DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)

DE PUTATIVE AMIDASE.

GN SCF37.03.

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxID=1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Oliver K., Harris D.;

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RX MEDLINE=97000351; PubMed=8843436;

RA Redenbach M., Kieser H.M., Denapait D., Eichner A., Cullum J.,

RA Kinashi H., Hopwood D.A.;

RT "A set of ordered cosmids and a detailed genetic and physical map for

RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL Mol. Microbiol. 21:77-96(1996).

DR EMBL; ALI33210; CAB61584.1; -

DR InterPro: IPR002502; -

DR Pfam; PF01510; Amidase2; 1.

SQ SEQUENCE 201 AA; 22749 MW; B8EF477E06A20468 CRC64;

Query Match

Best Local Similarity 85.7%; Pred. No. 16;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVPGSQH 7

Db 176 EVPGSDH 182

RESULT 9

Q9NPQ6

ID Q9NPQ6 PRELIMINARY; PRT; 192 AA.

DT 01-OCT-2000 (TRENBLrel. 15, Created)

DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)

DE LLPL, LCAT-LIKE LYOPHOSPHOLIPASE (FRAGMENT).

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Pluvinet R., Estivill X., Escarceller M., Sumoy L.;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Aufray C., Ansorge W., Ballabio A., Estivill X., Gibson K.,  
RA LeFrach H., Poustka A., Lundeberg J.;  
RT "The European IMAGE Consortium for integrated Molecular analysis of  
RT human gene transcripts."  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL389957; CAB97531.1; -.  
FT NON\_TER  
SQ SEQUENCE 192 AA; 21609 MW; 04A7AE8CB344F213 CRC64;

Query Match 84.6%; Score 33; DB 4; Length 192;  
Best Local Similarity 71.4%; Pred. No. 24;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EYPGSEQ 7  
Db 166 ELPGSEH 172  
I:|||||

RESULT 10  
Q9UG04 PRELIMINARY; PRT; 272 AA.  
AC Q9UG04;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE HYPOTHETICAL 31.0 KDA PROTEIN.  
GN DKF2P564A0122.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN:  
RA Wambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wienann S.;  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL110209; CAB53675.1; -.  
DR InterPro; IPR003386; -.  
DR Pfam; PF02450; LACT; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 272 AA; 31016 MW; ACC5E1680D7A720 CRC64;

Query Match 84.6%; Score 33; DB 4; Length 272;  
Best Local Similarity 71.4%; Pred. No. 35;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EYPGSEQ 7  
Db 246 ELPGSEH 252  
I:|||||

RESULT 11  
ID O80418 PRELIMINARY; PRT; 367 AA.  
AC O80418;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE NTC16 PROTEIN.  
GN NTC16.  
OS Nicotiana tabacum (Common tobacco).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
OC Solanales; Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Tanaka-Ueguchi M., Itoh H., Oyama N., Koshioka M., Matsuoka M.;  
RT "Over-expression of a tobacco homeobox gene, NTH15, decreases the  
RT expression of a gibberellin biosynthetic gene encoding GA 20-  
RT oxidase."  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB016084; BAA31690.1; -.  
DR Mendel; 31081; Nicta; 2972; 31081.  
DR InterPro; IPR002419; -.  
DR Pfam; PF00671; Fe\_Asc\_Oxidoreductase; 1.  
SQ SEQUENCE 367 AA; 42170 MW; 923BC90B3BBAC05 CRC64;

Query Match 84.6%; Score 33; DB 10; Length 367;  
Best Local Similarity 85.7%; Pred. No. 47;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 EYPGSEQ 7  
Db 149 EYPSSQH 155  
I:|||||

RESULT 12  
Q9Y2B3 PRELIMINARY; PRT; 412 AA.  
AC Q9Y2B3;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE LCAT-LIKE PROTEIN (LLPL).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99194552; PubMed=10092508;  
RA Taniyama Y., Shibata S., Kita S., Horikoshi K., Shirafuji H.,  
RA Sumino Y., Fujino M.;  
RT "Cloning and expression of a novel lysophospholipase which  
RT structurally resembles lecithin cholesterol acyltransferase."  
RL Biochem. Biophys. Res. Commun. 257:50-56(1999).  
DR EMBL; AB017494; BAA76877.1; -.  
DR InterPro; IPR003386; -.  
DR Pfam; PF02450; LACT; 1.  
SQ SEQUENCE 412 AA; 46657 MW; 1FEA8A5783AF050A CRC64;

Query Match 84.6%; Score 33; DB 4; Length 412;  
Best Local Similarity 71.4%; Pred. No. 53;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EYPGSEQ 7  
Db 386 ELPGSEH 392  
I:|||||

RESULT 13  
Q9VFE2 PRELIMINARY; PRT; 868 AA.  
AC Q9VFE2;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE CG3837 PROTEIN.  
GN CG3837.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Cocayne J.D.,  
 RA Ananadis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutcon G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA April J.F., Agbavani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RA "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 DR EMBL; AE003706; AAF5118.1; -.  
 DR Flybase; FBgn0038279; CG3837.  
 DR InterPro; IPR000494; -.  
 DR InterPro; IPR001777; -.  
 DR Pfam; PF01030; Recep\_Ldomain; 2.  
 DR SMART; SM00060; FN3; 1.  
 SQ SEQUENCE 868 AA; 98349 MW; F6562A64E72E7B21 CRC64;

Query Match 84.6%; Score 33; DB 5; Length 868;  
 Best Local Similarity 71.4%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYPGSOH 7  
 DB 812 ELPGTQH 818

RESULT 14  
 Q9FO18 PRELIMINARY; PRT; 95 AA.  
 AC Q9FO18:  
 DT 01-WAR-2001 (TREMBlrel. 16, Created)  
 DT 01-WAR-2001 (TREMBlrel. 16, Last sequence update)  
 DE REVERSE TRANSCRIPTASE-LIKE PROTEIN (FRAGMENT).  
 OS Amaranthus quitensis.  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;  
 OC Caryophyllales; Amaranthaceae; Amaranthus.  
 OX NCBI\_TaxID=107609;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TRANSPOS=COPIA-LIKE RETROTRANSPON;  
 RA He Y., Sun M.;  
 RT "Reverse transcriptase sequence evolution in copia-like  
 retrotransposons in plants."  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF232993; RAG44333.1; -.  
 FT NON\_TER 1  
 FT NON\_TER 95 95  
 SQ SEQUENCE 95 AA; 11392 MW; 8E931447E1683C5C CRC64;  
 Query Match 82.1%; Score 32; DB 10; Length 95;  
 Best Local Similarity 71.4%; Pred. No. 20;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYPGSOH 7  
 DB 26 EYPGKEH 32

RESULT 15  
 O70284 PRELIMINARY; PRT; 765 AA.  
 ID O70284:  
 AC O70284:  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-WAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE SINGLE-MINDED PROTEIN 1.  
 GN SIMI.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129SV;  
 RX MEDLINE=97020303; PubMed=8812055;  
 RA Fan C.M., Kuwana E., Bulfone A., Fletcher C.F., Copeland N.G.,  
 RA Jenkins N.A., Crews S., Martinez S., Puelles L., Rubenstein J.L.,  
 RA Tessier-Lavigne M.;  
 RT "Expression patterns of two murine homologs of Drosophila single-  
 minded suggest possible roles in embryonic patterning and in the  
 pathogenesis of Down syndrome."  
 RL Mol. Cell. Neurosci. 7:1-16(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129SV;  
 RA May N.R., Fan C.-M.;  
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF038857; AAC05481.1; -.  
 DR EMBL; AF044913; AAC05481.1; JOINED.  
 DR EMBL; AF038853; AAC05481.1; JOINED.  
 DR EMBL; AF038854; AAC05481.1; JOINED.  
 DR EMBL; AF038855; AAC05481.1; JOINED.  
 DR EMBL; AF038856; AAC05481.1; JOINED.  
 DR InterPro; IPR000014; -.  
 DR InterPro; IPR001092; -.  
 DR InterPro; IPR001610; -.  
 DR InterPro; IPR003015; -.  
 DR Pfam; PF00785; PAC; 1.  
 DR Pfam; PF00989; PAS; 2.  
 DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
 DR SMART; SM00086; PAC; 1.  
 SQ SEQUENCE 765 AA; 85575 MW; 46AC6BFD8A189126 CRC64;

Query Match 82.1%; Score 32; DB 11; Length 765;  
 Best Local Similarity 85.7%; Pred. No. 1.6e+02;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
| | | | |  
DB 423 ERPGSOH 429

RESULT 16

Q9FQJ8 PRELIMINARY; PRT; 95 AA.

AC Q9FQJ8;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE REVERSE TRANSCRIPTASE-LIKE PROTEIN (FRAGMENT).  
OS Amaranthus quitensis.  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales;  
OC Caryophyllales; Amaranthaceae; Amaranthus.  
OX NCBI\_TaxID=107609;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TRANSPOSON-COPIA-LIKE RETROTRANSPON;  
RA He Y., Sun M.;  
RT "Reverse transcriptase sequence evolution in copia-like  
retrotransposons in plants."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF232981; AAG44323.1; -;  
FT NON\_TER 1  
FT 95  
SQ SEQUENCE 95 AA; 11241 MW; 36403CBDC40D769A CRC64;

Query Match 79.5%; Score 31; DB 10; Length 95;  
Best Local Similarity 57.1%; Pred. No. 32;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
| | | | |  
DB 26 EIPGKEH 32

RESULT 17

Q9W2V0 PRELIMINARY; PRT; 122 AA.

AC Q9W2V0;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE CG15303 PROTEIN.  
GN CG15303.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Balwle R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bock J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan K.C., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RL "The genome sequence of Drosophila melanogaster."  
RT Science 287:2185-2195(2000).  
DR EMBL; AE003450; AAF46588.1; -;  
DR FlyBase; FBgn0030203; CG15303.  
SQ SEQUENCE 122 AA; 13139 MW; EF8D5468CFDE88E8 CRC64;

Query Match 79.5%; Score 31; DB 5; Length 122;  
Best Local Similarity 71.4%; Pred. No. 41;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
| | | | |  
DB 25 EAPGSRH 31

RESULT 18

Q9JL95 PRELIMINARY; PRT; 222 AA.

AC Q9JL95;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE EOSINOPHIL MAJOR BASIC PROTEIN 2.  
GN MBP2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=BONE MARROW;  
RA Macias M.P., Welch K.C., Denzler K.L., Larson K.A., Lee N.A.,  
RA Lee J.J.;  
RT "The Identification of a New Murine Eosinophil Major Basic Protein  
(mMBP) Gene: Cloning and Characterization of mMBP-2.";  
RL J. Leukoc. Biol. 0:0-0(2000).  
DR EMBL; AF202533; AAF26366.1; -;  
DR InterPro; IPR001304; -;  
DR InterPro; IPR002352; -;  
DR Pfam; PF00059; lectin.C.1;  
DR PRINTS; PR00770; EMAJORBASICP.  
DR PROSITE; PS00615; C-TYPE\_LECTIN\_1; 1.  
DR PROSITE; PS50041; C-TYPE\_LECTIN\_2; 1.  
DR SMART; SM00034; CLECT; 1.  
SQ SEQUENCE 222 AA; 25204 MW; 1A28C69888FC52FF CRC64;

Query Match 79.5%; Score 31; DB 11; Length 222;  
Best Local Similarity 85.7%; Pred. No. 75;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EYVGSQH 7

DB 61 EYVGSQH 67

RESULT 19

ID O86582 PRELIMINARY; PRT; 242 AA.

AC O86582;

DT 01-NOV-1998 (TrEMBLrel. 08, Created)

DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)

DE HYPOTHETICAL 26.0 KDA PROTEIN.

GN SC2A11.2IC.

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.

OX NCBI\_TaxID=1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Murphy L., Harris D.;

RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Parkhill J., Barrell B.G., Rajandream M.A.;

RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RX MEDLINE=97000351; PubMed=8843436;

RA Kadenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,

RA Kinashi H., Hopwood D.A.;

RT "A set of ordered cosmids and a detailed genetic and physical map for

the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL Mol. Microbiol. 21:77-96(1996).

DR EMBL; AL031184; CAA20190.1; -.

DR InterPro; IPR002502; -.

DR Pfam; PF01510; Amidase\_2; 1.

KW Hypothetical protein.

SQ SEQUENCE 242 AA; 26018 MW; 6DD10FF18A2EC544 CRC64;

Query Match

Best Local Similarity 79.5%; Score 31; DB 2; Length 242;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EYVGSQH 7

DB 205 EYVGTDH 211

RESULT 20

ID Q9Y63

AC Q9Y63 PRELIMINARY; PRT; 243 AA.

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

DE INTERLEUKIN-1 RECEPTOR ASSOCIATED PROTEIN KINASE (FRAGMENT).

GN ILIRAK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Reichwald K., Thiesen J., Wiehe T., Kioschis P., Straetling W.H.,

RA Rosenthal A., Platzer M.;

"Comparative analysis of the methyl CpG binding protein 2 locus in man

and mouse reveals new untranslated sequences.";

Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

EMBL; AF121351; AAF22115.1; -.

DR MGD; MGI:107420; Ilirak.

DR Pfam; PF00069; kinase; 1.

DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.

DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.

KW ATP-binding; Kinase; Receptor; Transferase.

FT NON\_TER 243 243

SQ SEQUENCE 243 AA; 26539 MW; ED22D000546F0E88 CRC64;

Query Match

Best Local Similarity 83.3%; Score 31; DB 11; Length 243;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VPGSQH 7

DB 12 VPGAQH 17

RESULT 21

ID Q9MA16

AC Q9MA16 PRELIMINARY; PRT; 260 AA.

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE F20B17.2.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

OC Brassicales; Brassicaceae; Arabidopsids.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,

RA Shinn P., Altafi H., Bei Q., Chin C., Chlou J., Choi E., Conn L.,

RA Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B., Lee J.,

RA Lenz C., Li J., Liu K., Liu S., Mukharsky N., Nguyen M.,

RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,

RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,

RA Ecker J.R.;

RT "Genomic sequence for Arabidopsis thaliana BAC F20B17 from chromosome

I.";

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Ecker J.R.;

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Ecker J.R.;

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,

RA Khan S., Kim C., Altafi H., Bei Q., Chin C., Chlou J., Choi E.,

RA Conn L., Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B.,

RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,

RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,

RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,

RA Theologis A., Ecker J.;

RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC010793; AAF68106.1; -.

SQ SEQUENCE 260 AA; 29128 MW; B149F22073AA0B92 CRC64;

Query Match

Best Local Similarity 79.5%; Score 31; DB 10; Length 260;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 VPGSQH 7

DB 12 VPGAQH 17

QY 1 EYPGSOH 7  
:|||||  
Db 66 KYPGSOH 72

RESULT 22  
Q9HNM5 PRELIMINARY; PRT; 274 AA.  
AC Q9HNM5  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE VNG2034H  
GN VNG2034H  
OS Halobacterium sp. (strain NRC-1)  
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
OC Halobacterium.  
OX NCBI\_TaxID=64091;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20504483; PubMed=11016950;  
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,  
RA Shukla H.D., Lasky S.R., Bailiga N.S., Thorsson V., Sbrogna J.,  
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,  
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;  
RT "Genome sequence of Halobacterium species NRC-1."  
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
DR EMBL; AE005096; AAG20195.1; -;  
SQ SEQUENCE 274 AA; 29944 MW; 2BE706911A76CD3D CRC64;

Query Match 79.5%; Score 31; DB 1; Length 274;  
Best Local Similarity 71.4%; Pred. No. 92;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
:|||||  
Db 117 EAPGSOH 123

RESULT 23  
O18391 PRELIMINARY; PRT; 331 AA.  
AC O18391  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-MAR-2001 (TREMBlrel. 05, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE PROBABLE SERINE HYDROLASE (EC 3.1.-.-) (KRAKEN PROTEIN).  
GN KRAKEN OR CG3943.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=EMBRYO;  
RX MEDLINE=99051329; PubMed=9831651;  
RA Chan H.Y.E., Harris S.J., O'Kane C.J.;  
RT "Identification and characterization of kraken, a gene encoding a  
putative hydrolytic enzyme in Drosophila melanogaster.";  
RL Gene 222:195-201(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Fabros B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Furlin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Haris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleeb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskaes R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
CC -!- FUNCTION: MAY HAVE A ROLE IN DETOXIFICATION AND DIGESTION DURING  
CC EMBRYOGENESIS AND LARVAL DEVELOPMENT.  
CC -!- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED BEFORE EMBRYONIC STAGE  
CC 11. AT STAGE 11, EXPRESSION IS CONCENTRATED IN THE FOREGUT AND  
CC POSTERIOR MIDGUT. BY STAGE 15, IN GASTRIC CAECAE, PHARYNX,  
CC POSTERIOR SPIRACLES AND ANTERIOR EDGE OF MIDGUT. AT THE END OF  
CC EMBRYOGENESIS, EXPRESSION IS CONFINED TO GASTRIC CAECAE. DURING  
CC THIRD INSTAR LARVAE, EXPRESSED AT LOW LEVELS IN GASTRIC CAECAE,  
CC MIDGUT AND HINDGUT AND HIGH LEVEL IN FAT BODY.  
CC -!- DEVELOPMENTAL STAGE: PROBABLY EXPRESSED BOTH MATERNALLY AND  
CC ZYGOTICALLY.  
CC -!- SIMILARITY: WEAK, TO FAMILY OF ESTERASES THAT GROUPS TOGETHER  
CC PSEUDOMONA TROPINESTERASE, DMPD, TDF AND XLF.  
CC EMBL; AF000516; CA04153.1; -;  
DR EMBL; AE003588; AAF51445.1; -;  
DR FlyBase; FBgn0020545; kraken.  
DR InterPro; IPR000073; -;  
DR InterPro; IPR000379; -;  
DR InterPro; IPR000734; -;  
DR Pfam; PF00561; abhydrolase; 1.  
DR PROSITE; PS00120; LIPASE\_SER; 1.  
KW Detoxification; Developmental protein; Digestion; Hydrolase;  
KW Serine esterase.  
FT ACT\_SITE 138 138 BY SIMILARITY.  
SQ SEQUENCE 331 AA; 37093 MW; 1FE39BC42AED4E69 CRC64;

Query Match 79.5%; Score 31; DB 5; Length 331;  
Best Local Similarity 71.4%; Pred. No. 1.1e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EYPGSOH 7  
:|||||  
Db 305 EYPGTHH 311

RESULT 24  
Q9RVJ7

ID Q9RVJ7 PRELIMINARY; PRT; 353 AA.  
 AC Q9RVJ7;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DE (S)-2-HYDROXY-ACID OXIDASE.  
 GN DR1031.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
 OX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R1;  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome sequence of the radioresistant bacterium Deinococcus  
 RT radiodurans R1.";  
 RL Science 286:1571-1577(1999).  
 DR EMBL: AF001954; AAF10604.1; -;  
 DR HSP; P05414; lGOX.  
 DR TIGR: DR1031; -;  
 DR InterPro; IPR000262; -;  
 DR InterPro; IPR003009; -;  
 DR Pfam; PF01070; FMN\_dh; 1.  
 DR PROSITE; PS00557; FMN\_HYDROXY\_ACID\_DH; 1.  
 SQ SEQUENCE 353 AA; 37877 MW; 14FB78FAE2E18C8D CRC64;

Query Match 79.5%; Score 31; DB 2; Length 353;  
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VPGSQH 7  
 DB 189 VPGSEH 194

RESULT 25  
 Q9RH15 PRELIMINARY; PRT; 371 AA.  
 AC Q9RH15;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DE HYPOTHETICAL 41.6 KDA PROTEIN.  
 OS Zymomonas mobilis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;  
 OC Zymomonas.  
 OX NCBI\_TaxID=542;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ZM4;  
 RA Lee J.S., Kang H.S.;  
 RT "Sequence analysis of 65G3 cosmid clone of Zymomonas mobilis ZM4  
 RT containing rna operon.";  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBDJ databases.  
 DR EMBL: AF088897; AAF18289.1; -;  
 DR InterPro; IPR000139; -;  
 DR ProDom; PD005242; -; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 371 AA; 41559 MW; 8D519E1509BAB64F CRC64;

Query Match 79.5%; Score 31; DB 2; Length 371;  
 Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VPGSQH 7  
 DB 288 LFGSQH 293

Search completed: July 16, 2001, 16:43:36  
 Job time: 470 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 16:36:18 ; Search time 30.3 Seconds  
(without alignments)  
13.962 Million cell updates/sec

Title: US-09-786-648-5  
Perfect score: 107  
Sequence: 1 GETQVEVPGSHDSQKKAI 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	93	2	US-08-292-968-26
2	107	100.0	93	2	US-08-467-974-26
3	107	100.0	93	2	US-08-467-536-26
4	107	100.0	93	3	US-08-467-976-26
5	107	100.0	93	4	US-09-082-514-26
6	107	100.0	371	2	US-08-829-026A-6
7	101	94.4	102	3	US-08-952-337-5
8	101	94.4	102	3	US-08-952-337-6
9	101	94.4	103	2	US-08-472-171-2
10	101	94.4	103	2	US-08-894-526-2
11	101	94.4	103	2	US-09-013-047-2
12	101	94.4	103	4	US-09-374-597-2
13	101	94.4	103	4	US-09-191-852-21
14	101	94.4	103	5	PCT-US95-13376-21
15	101	94.4	123	3	US-08-952-337-1
16	101	94.4	123	3	US-08-952-337-2
17	101	94.4	124	2	US-08-747-410-2
18	89	83.2	124	1	US-08-449-045C-4
19	89	83.2	124	2	US-08-435-605A-12
20	89	83.2	124	6	5223610-3
21	54	50.5	448	4	US-08-878-989-2
22	54	50.5	448	4	US-09-272-796-2
23	43	40.2	1810	5	PCT-US95-11684-4
24	42	39.3	855	2	US-09-027-337-2
25	41	38.3	1022	1	US-08-271-364A-8
26	41	38.3	1022	2	US-08-222-715B-27
27	40.5	37.9	856	3	US-08-709-784-2

28	40.5	37.9	862	2	US-08-209-521-23	Sequence 23, Appl
29	40.5	37.9	862	2	US-08-209-521-30	Sequence 30, Appl
30	40.5	37.9	862	4	US-09-059-461-2	Sequence 2, Appl1
31	40.5	37.9	862	4	US-08-961-810-133	Sequence 133, App
32	40.5	37.9	862	4	US-08-352-902D-133	Sequence 133, App
33	40	37.4	774	3	US-08-902-632-2	Sequence 2, Appl1
34	40	37.4	774	3	US-09-073-354-1	Sequence 1, Appl1
35	40	37.4	774	3	US-08-556-005A-1	Sequence 1, Appl1
36	40	37.4	774	4	US-09-073-259-1	Sequence 1, Appl1
37	40	37.4	774	4	US-09-363-095-1	Sequence 1, Appl1
38	40	37.4	774	4	US-09-418-027-1	Sequence 1, Appl1
39	40	37.4	778	2	US-08-906-925-4	Sequence 12, Appl
40	40	37.4	779	1	US-08-375-134-12	Sequence 12, Appl
41	40	37.4	779	5	PCT-US95-15263-12	Sequence 12, Appl
42	39	36.4	75	1	US-08-350-884-35	Sequence 35, Appl
43	39	36.4	75	2	US-08-709-173-35	Sequence 35, Appl
44	39	36.4	75	2	US-08-709-177-35	Sequence 35, Appl
45	39	36.4	75	2	US-08-833-678A-3	Sequence 3, Appl1

## ALIGNMENTS

RESULT 1  
US-08-292-968-26  
; Sequence 26, Application US/08292968  
; Patent No. 5856122  
; GENERAL INFORMATION:  
; APPLICANT: READ, Randy J.  
; APPLICANT: STEIN, Penelope E.  
; APPLICANT: COCKLE, Stephen A.  
; APPLICANT: OMEN, Raymond P.  
; APPLICANT: LOOSMORE, Sheena  
; APPLICANT: KLEIN, Michel H.  
; APPLICANT: ARMSTRONG, Glen D.  
; APPLICANT: HAYES, Bart  
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/292,968  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/110,947  
; FILING DATE: 24-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/251,121  
; FILING DATE: 31-MAY-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-388  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 93 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-08-292-968-26

Query Match 100.0%; Score 107; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 2.2e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHDSOKKAI 21  
|||||

DB 35 GETFOVEVPGSQHDSOKKAI 55

RESULT 2

US-08-467-974-26  
; Sequence 26, Application US/08467974

; Patent No. 5965385

; GENERAL INFORMATION:

; APPLICANT: READ, Randy J.

; APPLICANT: STEIN, Penelope E.

; APPLICANT: COCKLE, Stephen A.

; APPLICANT: COHEN, Raymond P.

; APPLICANT: LOOSMORE, Sheena

; APPLICANT: KLEIN, Michel H.

; APPLICANT: ARMSTRONG, Glen D.

; APPLICANT: HAZES, Bart

; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN

; NUMBER OF SEQUENCES: 46

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: Suite 701, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/467,974

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/467,536

; FILING DATE: 06-JUN-1995

; APPLICATION NUMBER: US 08/292,968

; FILING DATE: 22-AUG-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/251,121

; FILING DATE: 31-MAY-1994

; APPLICATION DATA:

; APPLICATION NUMBER: US 08/110,947

; FILING DATE: 24-AUG-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: STEWART, Michael I.

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-454 MIS:vg

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1163

; INFORMATION FOR SEQ ID NO: 26:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 93 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-467-974-26

Query Match 100.0%; Score 107; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 2.2e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHDSOKKAI 21

|||||

DB 35 GETFOVEVPGSQHDSOKKAI 55

RESULT 4

US-08-467-976-26

QY 1 GETFOVEVPGSQHDSOKKAI 21  
|||||

DB 35 GETFOVEVPGSQHDSOKKAI 55

RESULT 3

US-08-467-536-26

; Sequence 26, Application US/08467536

; Patent No. 5977304

; GENERAL INFORMATION:

; APPLICANT: READ, Randy J.

; APPLICANT: STEIN, Penelope E.

; APPLICANT: COCKLE, Stephen A.

; APPLICANT: COHEN, Raymond P.

; APPLICANT: LOOSMORE, Sheena

; APPLICANT: KLEIN, Michel H.

; APPLICANT: ARMSTRONG, Glen D.

; APPLICANT: HAZES, Bart

; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN

; NUMBER OF SEQUENCES: 46

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: Suite 701, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/467,536

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/292,968

; FILING DATE: 22-AUG-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/251,121

; FILING DATE: 31-MAY-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/110,947

; FILING DATE: 24-AUG-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: STEWART, Michael I.

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-455 MIS:vg

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1163

; INFORMATION FOR SEQ ID NO: 26:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 93 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-467-536-26

Query Match 100.0%; Score 107; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 2.2e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHDSOKKAI 21

|||||

DB 35 GETFOVEVPGSQHDSOKKAI 55

RESULT 4

US-08-467-976-26

```

; Sequence 26, Application US/08467976
; Patent No. 6018022
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J.
; APPLICANT: STEIN, Penelope E.
; APPLICANT: COCKLE, Stephen A.
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: KLEIN, Michel H.
; APPLICANT: ARMSTRONG, Glen D.
; APPLICANT: HAZES, Bart
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,976
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,968
; FILING DATE: 22-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/251,121
; FILING DATE: 31-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/110,947
; FILING DATE: 24-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-453 MIS.vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-467-976-26

```

```

Query Match 100.0%; Score 107; DB 3; Length 93;
Best Local Similarity 100.0%; Pred. No. 2.2e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GETFQVEVPGSQHDSQKKAI 21
Db 35 GETFQVEVPGSQHDSQKKAI 55

```

```

RESULT 5
US-09-514-26
; Sequence 26, Application US/09082514
; Patent No. 6168928
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J.
; APPLICANT: STEIN, Penelope E.
; APPLICANT: COCKLE, Stephen A.
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.

```

```

; APPLICANT: ARMSTRONG, Glen D.
; APPLICANT: HAZES, Bart
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/082,514
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,968
; FILING DATE: 24-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-810
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-082-514-26

Query Match 100.0%; Score 107; DB 4; Length 93;
Best Local Similarity 100.0%; Pred. No. 2.2e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFQVEVPGSQHDSQKKAI 21
Db 35 GETFQVEVPGSQHDSQKKAI 55

RESULT 6
US-08-829-026A-6
; Sequence 6, Application US/08829026A
; Patent No. 5837825
; GENERAL INFORMATION:
; APPLICANT: Meinersmann, Richard J.
; APPLICANT: Khoury, Christian A.
; TITLE OF INVENTION: Campylobacter Jejuni Flagellin-Escherichia Coli LT-B Fusion
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Janelle S. Graeter
; STREET: Room 411, Building 005, BARC-W
; CITY: Beltsville
; STATE: MD
; COUNTRY: USA
; ZIP: 20705
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/829,026A
; FILING DATE: 18-AUG-1997
; CLASSIFICATION: 435

```

ATTORNEY/AGENT INFORMATION:  
NAME: Graeter, Janelle S.  
REGISTRATION NUMBER: 35,024  
REFERENCE/DOCKET NUMBER: 0106.97  
TELEPHONE: 301-504-5676  
TELEFAX: 301-504-5060  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 371 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-829-026A-6

Query Match 100.0%; Score 107; DB 2; Length 371;  
Best Local Similarity 100.0%; Pred. No. 1.2e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSOHIDSOKKAI 21  
Db 49 GETFOVEVPGSOHIDSOKKAI 69

RESULT 7  
US-08-952-337-5  
Sequence 5, Application US/08952337  
Patent No. 6019973  
GENERAL INFORMATION:  
APPLICANT: Holmgren, Jan  
APPLICANT: Lebens, Michael R.  
TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE  
FILE REFERENCE: 3846/0D758  
CURRENT APPLICATION NUMBER: US/08/952.337  
CURRENT FILING DATE: 1998-01-05  
EARLIER APPLICATION NUMBER: PCT/SE96/00570  
EARLIER FILING DATE: 1996-05-02  
EARLIER APPLICATION NUMBER: SE 9501682-0  
EARLIER FILING DATE: 1995-05-05  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 5  
LENGTH: 102  
TYPE: PRT  
ORGANISM: Vibrio cholerae  
US-08-952-337-5

Query Match 94.4%; Score 101; DB 3; Length 102;  
Best Local Similarity 95.2%; Pred. No. 2.7e-10;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSOHIDSOKKAI 21  
Db 44 GATFOVEVPGSOHIDSOKKAI 64

RESULT 8  
US-08-952-337-6  
Sequence 6, Application US/08952337  
Patent No. 6019973  
GENERAL INFORMATION:  
APPLICANT: Holmgren, Jan  
APPLICANT: Lebens, Michael R.  
TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE  
FILE REFERENCE: 3846/0D758  
CURRENT APPLICATION NUMBER: US/08/952.337  
CURRENT FILING DATE: 1998-01-05  
EARLIER APPLICATION NUMBER: PCT/SE96/00570  
EARLIER FILING DATE: 1996-05-02

EARLIER APPLICATION NUMBER: SE 9501682-0  
EARLIER FILING DATE: 1995-05-05  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 6  
LENGTH: 102  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-08-952-337-6

Query Match 94.4%; Score 101; DB 3; Length 102;  
Best Local Similarity 95.2%; Pred. No. 2.7e-10;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSOHIDSOKKAI 21  
Db 44 GATFOVEVPGSOHIDSOKKAI 64

RESULT 9  
US-08-472-171-2  
Sequence 2, Application US/08472171  
Patent No. 5932714  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M.  
APPLICANT: Yacoub, Reza K.  
APPLICANT: Zealey, Gavin R.  
APPLICANT: Klein, Michel H.  
TITLE OF INVENTION: Expression Of Gene Products From  
TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 330 University Avenue, Suite 701  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472.171  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/393.334  
FILING DATE: 23-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-507 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 416-595-1155  
TELEFAX: 416-595-1163  
TELEX: 065-24567 Simbas  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-472-171-2

Query Match 94.4%; Score 101; DB 2; Length 103;  
Best Local Similarity 95.2%; Pred. No. 2.8e-10;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSOHIDSOKKAI 21

```

STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/013,047
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/472,171
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/393,334
FILING DATE: 23-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-507 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: 416-595-1155
TELEFAX: 416-595-1163
TELEX: 065-24567 Simbas
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-013-047-2

Query Match          94.4%; Score 101; DB 2; Length 103;
Best Local Similarity 95.2%; Pred. No. 2.8e-10;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHIDSOKKAI 21
DB 45 GATFOVEVPGSQHIDSOKKAI 65

RESULT 12
US-09-374-597-2
; Sequence 2, Application US/09374597
; Patent No. 6140082
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yacoub, Reza K.
; APPLICANT: Zealey, Gavin R.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Expression Of Gene Products From
; TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 330 University Avenue, 6th Floor
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/374,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```



ZIP: 55402  
COMPUTER READABLE FORM:

```

: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/449,045C
: FILING DATE: 24-MAY-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/694,733
: FILING DATE: 02-MAY-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/271,222

```





CLONE: 40194  
US-08-878-989-2

Query Match 50.5%; Score 54; DB 2; Length 448;  
Best Local Similarity 50.0%; Pred. No. 0.21;  
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHDSOKKA 20  
::: | | | | | | | | | |

Db 269 QGSINVHTPNSQKVDOSKAA 288

RESULT 22

US-09-272-796-2

; Sequence 2, Application US/09272796

; Patent No. 6207148

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Corley, Neil C.

; APPLICANT: Guegler, Karl G.

; APPLICANT: Lal, Preeti

; APPLICANT: Goli, Surya K.

; APPLICANT: Shah, Purvi

; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN

; TITLE OF INVENTION: KINASES

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/272,796

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/878,989

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J J

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0321 US

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 448 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: TBLINOT01

; CLONE: 40194

US-09-272-796-2

Query Match 50.5%; Score 54; DB 4; Length 448;  
Best Local Similarity 50.0%; Pred. No. 0.21;  
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHDSOKKA 20  
::: | | | | | | | | | |

Db 269 QGSINVHTPNSQKVDOSKAA 288

RESULT 23

PCT-US95-11684-4

; Sequence 4, Application PC/TUS9511684

; GENERAL INFORMATION:

; APPLICANT: THE SCRIPPS RESEARCH INSTITUTE

; TITLE OF INVENTION: CYTOTOXIC DERIVATIVES THAT STIMULATE

; TITLE OF INVENTION: ATTACHMENT AND NEURITE OUTGROWTH, AND METHODS OF MAKING

; NUMBER OF SEQUENCES: 28

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: The Scripps Research Institute, Office of

; ADDRESSEE: Patent Counsel

; STREET: 10666 North Torrey Pines Road, TPC 8

; CITY: La Jolla

; STATE: CA

; COUNTRY: USA

; ZIP: 92037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/11684

; FILING DATE: 14-SEP-1995

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/308,359

; FILING DATE: 16-SEP-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Logan, April C.

; REGISTRATION NUMBER: 33,950

; REFERENCE/DOCKET NUMBER: BEC0019P

; TELEPHONE: 619-554-2937

; TELEFAX: 619-554-6312

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1810 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

PCT-US95-11684-4

Query Match 40.2%; Score 43; DB 5; Length 1810;

Best Local Similarity 58.3%; Pred. No. 90;

Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 ETFOVEVPGSQH 13

::: | | | | |

Db 1089 ETWNITVPGGQH 1100

RESULT 24

US-09-027-337-2

; Sequence 2, Application US/09027337B

; Patent No. 5972616

; GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.

; APPLICANT: Tanimoto, Hirotsu

; TITLE OF INVENTION: TAG-15: An Extracellular Serine Protease Overexpressed in

; TITLE OF INVENTION: Breast and Ovarian Carcinomas

; FILE REFERENCE: D6064

; CURRENT APPLICATION NUMBER: US/09/027,337B

; CURRENT FILING DATE: 1998-02-20

; NUMBER OF SEQ ID NOS: 13

; SEQ ID NO 2

; LENGTH: 855

; TYPE: PRT

ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Amino acid sequence of TADG-15 encoded by nucleotides  
OTHER INFORMATION: 23 to 2589 of Sequence 1  
Patent No. 5972616  
US-09-027-337-2

Query Match 39.3%; Score 42; DB 2; Length 855;  
Best Local Similarity 50.0%; Pred. No. 54;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
QY 3 TFQVEVPGSQHI 14  
DB 367 TWNIEVPNNQHV 378  
I: :||| :||:  
|: :||| :||:

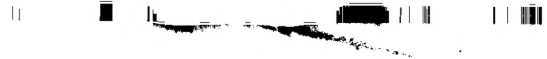
RESULT 25  
US-08-271-364A-8  
; Sequence 8, Application US/08271364A  
; Patent No. 5756334  
; GENERAL INFORMATION:  
; APPLICANT: PERLER, FRANCINE B.  
; APPLICANT: SOUTHWORTH, MAURICE W.  
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE DNA POLYMERASE  
; TITLE OF INVENTION: FROM ARCHAEABACTERIA  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NEW ENGLAND BIOLABS, INC.  
; STREET: 32 TOZER ROAD  
; CITY: BEVERLY  
; STATE: MASSACHUSETTS  
; COUNTRY: US  
; ZIP: 01915  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/271,364A  
; FILING DATE: 06-JUL-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/811,421  
; FILING DATE: 18-DEC-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/686,340  
; FILING DATE: 17-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/626,057  
; FILING DATE: 11-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/513,994  
; FILING DATE: 26-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WILLIAMS, GREGORY D.  
; REGISTRATION NUMBER: 30901  
; REFERENCE/DOCKET NUMBER: NEB-101  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (508) 927-5054  
; TELEFAX: (508) 927-1705  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1022 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-271-364A-8

Query Match 38.3%; Score 41; DB 1; Length 1022;

Best Local Similarity 46.7%; Pred. No. 1e+02;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
QY 1 GETFOVEVPGSQHID 15  
DB 247 GDSFAVEIKGRIFHD 261  
I: :||| :||:  
|: :||| :||:

Search completed: July 16, 2001, 16:36:18  
Job time: 207 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 16:37:02 ; Search time 34.9 Seconds  
(without alignments)  
45.836 Million cell updates/sec

Title: US-09-786-648-5

Perfect score: 107

Sequence: 1 GETFQVEVPGSQHDSQKAI 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	124	1 QLECB	heat-labile entero
2	95	88.8	124	1 XVVCB	cholera enterotoxi
3	49	45.8	255	2 A86457	hypothetical prote
4	47	43.9	392	2 T04150	RAD23 protein homo
5	46	43.0	91	2 C96580	hypothetical prote
6	46	43.0	574	2 C86400	hypothetical prote
7	45	42.1	439	2 T49189	kinesin heavy chal
8	45	42.1	2733	2 S15760	genome polyprotein
9	44	41.1	263	2 S78364	conserved hypotet
10	44	41.1	374	2 T19866	hypothetical prote
11	44	41.1	864	2 T49574	probable carnitine
12	44	41.1	2731	1 VFTHJH	genome polyprotein
13	43.5	40.7	641	2 C84726	probable receptor-
14	43	40.2	128	1 UQDOR	ubiquitin / riboso
15	43	40.2	154	1 UQDOR7	ubiquitin 18 - sli
16	43	40.2	228	2 D34080	ubiquitin (clone 1
17	43	40.2	229	2 B27806	hypothetical prote
18	43	40.2	368	2 T04861	polyubiquitin 5 (c
19	43	40.2	380	2 C34080	polyubiquitin 5 (c
20	43	40.2	380	2 B34080	polyubiquitin 5 (c
21	43	40.2	381	2 A27806	polyubiquitin 7 (c
22	43	40.2	532	2 A34080	tenascin precursor
23	43	40.2	1810	1 A32230	probable peptide c
24	42	39.3	204	2 E82963	stress-activated p
25	42	39.3	427	2 JC5694	hypothetical prote
26	42	39.3	1742	2 S76110	chorismate synthas
27	41.5	38.8	376	1 S17246	4-aminobutyrate tr
28	41.5	38.8	500	2 JCA022	probable peptide c
29	41	38.3	166	2 E85512	

30	41	38.3	166	2 E64748	translation releas
31	41	38.3	260	2 C96827	protein F20B17.2 [
32	41	38.3	269	1 A25973	pertussis toxin ch
33	41	38.3	386	1 S66056	yaaN protein - Bac
34	41	38.3	399	2 T46898	queuine tRNA-ribos
35	41	38.3	550	2 T01770	hypothetical prote
36	41	38.3	683	2 S01433	repressor protein
37	41	38.3	787	2 G81692	inner membrane pro
38	41	38.3	1150	2 A83978	pyruvate carboxyla
39	41	38.3	1167	2 T34020	zinc finger protei
40	41	38.3	1702	2 S42459	DNA-directed DNA p
41	41	38.3	2911	2 T20566	hypothetical prote
42	40.5	37.9	374	2 D81715	conserved hypotet
43	40.5	37.9	862	2 S47598	mutl protein homol
44	40	37.4	148	2 H72513	translation initia
45	40	37.4	149	1 UQNCR	ubiquitin / riboso

#### ALIGNMENTS

RESULT 1

QLECB  
heat-labile enterotoxin chain B precursor - Escherichia coli  
C:Species: Escherichia coli  
C:Date: 29-Jun-1981 #sequence\_revision 29-Jun-1981 #text\_change 18-Jun-1999  
C:Accession: A01820; B26946; I41194; I41287; I67644; A61475  
R:Dallas, W.S.; Falkow, S.  
Nature 288, 499-501, 1980  
A:Title: Amino acid sequence homology between cholera toxin and Escherichia coli heat  
A:Reference number: A01820; MUID:81074965  
A:Accession: A01820  
A:Molecule type: mRNA  
A:Residues: 1-124 <DAL>  
R:Yamamoto, T.; Gojobori, T.; Yokota, T.  
J. Bacteriol. 169, 1352-1357, 1987  
A:Title: Evolutionary origin of pathogenic determinants in enterotoxigenic Escherich  
A:Reference number: A26946; MUID:87137303  
A:Accession: B26946  
A:Molecule type: DNA  
A:Residues: 1-27, 'E', 29-63, 'K', 65-124 <YAM>  
A:Cross-references: EMBL:M15363; NID:gl48335; PIDN:AAA24792.1; PID:gl48336  
R:Leong, J.; Vinal, A.C.; Dallas, W.S.  
Infect. Immun. 48, 73-77, 1985  
A:Title: Nucleotide sequence comparison between heat-labile toxin B-subunit cistrons  
A:Reference number: I41194; MUID:85156481  
A:Accession: I41194  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-5, 'F', 7-17, 'C', 19-24, 'S', 26-27, 'E', 29-33, 'H', 35-63, 'K', 65-66, 'A', 68-122  
A:Cross-references: GB:M17874; NID:gl45830; PIDN:AAA98064.1; PID:gl45831  
A:Experimental source: plasmid ENT-R PCG86  
R:Ibrahim, I.; Gentz, R.  
J. Biol. Chem. 262, 10189-10194, 1987  
A:Title: A functional interaction between the signal peptide and the translation appa  
ticulum.  
A:Reference number: I41287; MUID:87280041  
A:Accession: I41287  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-22 <RE2>  
A:Cross-references: GB:M17101; NID:gl46375; PIDN:AAA23973.1; PID:gl46376  
R:Inoue, T.; Tsuji, T.; Koto, M.; Imamura, S.; Miyama, A.  
FEMS Microbiol. Lett. 108, 157-161, 1993  
A:Title: Amino acid sequence of heat-labile enterotoxin from chicken enterotoxigenic  
A:Reference number: I53542; MUID:93252225  
A:Accession: I67644  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-17, 'C', 19, 'Y', 21-24, 'S', 26-27, 'E', 29-63, 'K', 65-66, 'A', 68-122, 'E', 124 <R  
A:Cross-references: GB:S60731; NID:gl408994; PIDN:AAC60441.1; PID:gl408996  
R:Tsuji, T.; Lida, T.; Honda, T.; Miwatani, T.; Nagahama, M.; Sakurai, J.; Wada, K.;  
Microb. Pathog. 2, 381-390, 1987

A:Title: A unique amino acid sequence of the B subunit of a heat-labile enterotoxin isolated from *Vibrio cholerae* O139.  
A:Reference number: A61475; MUID:89180953  
A:Accession: A61475  
A:Molecule type: protein  
A:Residues: 22-24, 'S', 26-27, 'E', 29-63, 'K', 65-66, 'A', 68-95, 'A', 97-122, 'E', 124 <TSU>  
A:Experimental source: strain 240-3  
C:Complex: the heat-labile enterotoxin molecule contains one A chain and five or six B chains  
C:Function:  
A:Description: the biological activity of the toxin is produced by the A chain, which acts as a superfamily: cholera enterotoxin beta chain  
C:Superfamily: cholera enterotoxin beta chain  
C:Keywords: enterotoxin  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-124/Product: heat-labile enterotoxin chain B #status predicted <MAT>  
F:30-107/Disulfide bonds: #status predicted

Query Match 100.0%; Score 107; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 4.6e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHDSQKAI 21  
|||||  
DB 66 GETFOVEVPGSQHDSQKAI 86

RESULT 2  
XVVCB  
cholera enterotoxin, B chain precursor VC1456 [validated] - Vibrio cholerae (strain N169)  
N:Alternate names: enterotoxin beta chain  
C:Species: Vibrio cholerae  
C:Date: 24-Apr-1984 #sequence\_revision 01-Sep-2000 #text\_change 02-Feb-2001  
C:Accession: S14624; S39238; S39241; H82196; JCI1078; S17666; PC1010; A05130; A01819; A3819  
R:Dams, E.; de Wolf, M.; Dierick, W.  
submitted to the EMBL Data Library, March 1991  
A:Description: Correction of the cholera toxin nucleotide sequence of the Vibrio cholerae  
A:Reference number: S14624  
A:Accession: S14624  
A:Molecule type: DNA  
A:Residues: 1-124 <DM>  
A:Cross-references: EMBL:X59786; NID:948420; PIDN:CAA1593.1; PID:948422  
A:Experimental source: strain 2125  
R:Lebens, M.; Holmgren, J.  
submitted to the EMBL Data Library, November 1993  
A:Description: Structure and arrangement of the Cholera toxin genes in vibrio Cholera O1  
A:Reference number: S39238  
A:Accession: S39238  
A:Molecule type: DNA  
A:Residues: 1-124 <LEB>  
A:Cross-references: EMBL:X76390; NID:9433856; PIDN:CAA53973.1; PID:9433857  
A:Accession: S39241  
A:Molecule type: DNA  
A:Residues: 1-124 <LEW>  
A:Cross-references: EMBL:X76391; NID:9433859; PIDN:CAA53976.1; PID:9433861  
R:Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, R.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833  
A:Accession: H82196  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-124 <HEI>  
A:Cross-references: GB:AE004224; GB:AE003852; NID:96555952; PIDN:AAF94613.1; GSPDB:GN001  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
R:Shi, C.H.; Cao, C.; Zhang, J.S.; Ma, Q.J.  
Chinese Biochem. J. 9, 395-399, 1993  
A:Title: Nucleotide sequence analysis of the gene encoding the classical biotype V.cholerae  
A:Reference number: JCI1078  
A:Accession: JCI1078  
A:Molecule type: DNA  
A:Residues: 1-20, 'Q', 22-31, 'Q', 33-38, 'H', 40-49, 'G', 51-67, 'T', 69-124 <SHI>  
A:Experimental source: classical biotype strain 569B

R:Dams, E.; de Wolf, M.; Dierick, W.  
Biochim. Biophys. Acta 1050, 139-141, 1991  
A:Title: Nucleotide sequence analysis of the CT operon of the Vibrio cholerae classic  
A:Reference number: S17665; MUID:91355224  
A:Accession: S17666  
A:Molecule type: DNA  
A:Residues: 1-38, 'H', 40-67, 'T', 69-124 <DA2>  
A:Cross-references: EMBL:X58785; NID:948888; PIDN:CAA41591.1; PID:948890  
R:Ma, Q.J.; Liu, C.X.; Xiong, L.S.; Yu, X.Q.  
Sci. Sin. B Chem. Biol. Agric. Med. Earth Sci. 34, 274-280, 1991  
A:Title: B subunit of cholera toxin produced in Escherichia coli.  
A:Reference number: PC1010  
A:Accession: PC1010  
A:Molecule type: protein  
A:Residues: 22-38, 'H', 40-41 <MAQ>  
R:Mekalanos, J.J.; Swartz, D.J.; Pearson, G.D.N.; Harford, N.; Groyne, F.; de Wilde, F.  
Nature 306, 551-557, 1983  
A:Reference number: A93320; MUID:84068199  
A:Accession: A05130  
A:Molecule type: DNA  
A:Residues: 1-32, 'S', 34-74, 'S', 76-124 <MEK>  
A:Cross-references: GB:X00171; NID:948347; PIDN:CAA24996.1; PID:9758351  
R:Kurosky, A.; Markel, D.E.; Peterson, J.W.  
J. Biol. Chem. 252, 7257-7264, 1977  
A:Title: Covalent structure of the beta chain of cholera enterotoxin.  
A:Reference number: A01819; MUID:78005537  
A:Accession: A01819  
A:Molecule type: protein  
A:Residues: 22-38, 'H', 40-42, 'N', 44-67, 'T', 69-90, 'N', 92-124 <KUR>  
R:Lai, C.Y.  
J. Biol. Chem. 252, 7249-7256, 1977  
A:Title: Determination of the primary structure of cholera toxin B subunit.  
A:Reference number: A38033; MUID:78005536  
A:Accession: A38033  
A:Molecule type: protein  
A:Residues: 22-38, 'H', 40-42, 'N', 44-67, 'T', 69, 'E', 71-90, 'N', 92-124 <LAI>  
A:Note: the difference at residue 70 may be due to deamidation during preparation  
R:Nakashima, Y.; Napiorkowski, P.; Schafer, D.E.; Konigsberg, W.H.  
FEBS Lett. 88, 275-278, 1976  
A:Title: Primary structure of the B subunit of cholera enterotoxin.  
A:Reference number: A38034; MUID:77026365  
A:Accession: A38034  
A:Molecule type: protein  
A:Residues: 22-38, 'H', 40-67, 'T', 69, 'E', 71, 'QS', 74-75, 'VE', 78-86, 'Q', 88-99, 'Q', 101-103  
R:Rakao, I.; Watanabe, H.; Shimonishi, Y.  
Eur. J. Biochem. 146, 503-508, 1985  
A:Title: Facile identification of protein sequences by mass spectrometry.  
A:Reference number: A21910; MUID:85126976  
A:Accession: A21910  
A:Molecule type: protein  
A:Residues: 22-38, 'H', 40-42, 'N', 44-67, 'T', 69-90, 'N', 92-124 <TAK>  
A:Experimental source: biotype inaba 569B  
A:Note: Asn-65 was partially deaminated to Asp  
C:Comment: The authors translated the codon TCA for residue 33 as Tyr.  
C:Genetics:  
A:Gene: VC1456  
A:Map position: 1  
C:Complex: the cholera enterotoxin molecule contains three kinds of chains; an alpha chain noncovalently with the subunit B, an aggregate of five beta chains  
C:Function:  
A:Description: involved in binding of the toxin to cell membranes  
C:Superfamily: cholera enterotoxin beta chain  
C:Keywords: enterotoxin; toxin  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-124/Product: cholera enterotoxin chain B #status predicted <MAT>  
F:30-107/Disulfide bonds: #status experimental

Query Match 88.8%; Score 95; DB 1; Length 124;  
Best Local Similarity 90.5%; Pred. No. 4e-08;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHDSQKAI 21

```
Db 66 CAIFQVEVPGSQHDSQKAI 86
|||||
RESULT 3
A86457
hypothetical protein AAG21605.1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: A86457
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: A86457
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-255 <STO>
A:CROSS-references: GB:AE005172; NID:g10645493; PIDN:AAG21605.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 45.8%; Score 49; DB 2; Length 255;
Best Local Similarity 45.0%; Pred. No. 2.4;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ETFOVEVPGSQHDSQKAI 21
|||||
Db 101 ETRFVSGPGQHRNKRDSAV 120
|||||

RESULT 4
T04150
RAD23 protein homolog - rice
C:Species: Oryza sativa (rice)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: T04150
R:Schultz, T.F.; Quatrano, R.S.
Plant Mol. Biol. 34, 557-562, 1997
A:Title: Characterization and expression of a rice RAD23 gene.
A:Reference number: Z08695; MUID:97369378
A:Accession: T04150
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-392 <SCH>
A:CROSS-references: EMBL:U63530; NID:g1488296; PIDN:AAB65841.1; PID:g1488297
C:Genetics:
A:Gene: RAD23
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 43.9%; Score 47; DB 2; Length 392;
Best Local Similarity 42.9%; Pred. No. 8.1;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GETFOVEVPGSQHDSQKAI 21
|||||
Db 10 GSTFQIEVDSAQKADVVKRII 30
|||||

RESULT 5
C96580
hypothetical protein F1511.6 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
```

```
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C96580
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C96580
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-91 <STO>
A:CROSS-references: GB:AE005173; NID:g4587538; PIDN:AAD25769.1; GSPDB:GN00141
C:Genetics:
A:Gene: F1511.6
A:Map position: 1

Query Match 43.0%; Score 46; DB 2; Length 91;
Best Local Similarity 47.8%; Pred. No. 2.5;
Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GETFOVEVPGSQHDSQKAI 21
|||||
Db 58 GKTFFLEVKGSEIIQVKNMI 78
|||||

RESULT 6
C86400
hypothetical protein AAD45990.1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C86400
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C86400
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-574 <STO>
A:CROSS-references: GB:AE005172; NID:g5668763; PIDN:AAD45990.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 43.0%; Score 46; DB 2; Length 574;
Best Local Similarity 58.8%; Pred. No. 18;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 TFOVEVPGSQHDSQKAI 19
|||||
Db 19 TFFVDPSSQHIKVKKK 35
|||||

RESULT 7
T49189
kinesin heavy chain-like protein - Arabidopsis thaliana
N:Alternate names: protein MAA21.110
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
```

C:Accession: T49189  
R:Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Mewes, H.W.; Rudd, S.; Lemcke, K.  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z25018  
A:Accession: T49189  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-439 <RIE>  
A:Cross-references: EMBL:AL163818; GSPDB:GN00061; ATSP:MAA21.110  
A:Experimental source: cultivar Columbia; BAC clone MAA21  
C:Genetics:  
A:Gene: ATSP:MAA21.110  
A:Map position: 3  
A:Introns: 39/3; 74/1; 97/3; 138/3; 147/2; 193/3; 200/2; 273/3; 304/2; 341/3; 401/3

Query Match 42.1%; Score 45; DB 2; Length 439;  
Best Local Similarity 38.1%; Pred. No. 19;  
Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHIDSOKKAI 21  
|||:|||||:  
Db 91 GKTYSMEGPGIQDCDEHNKGL 111

RESULT 8  
S15760  
genome polyprotein - murine hepatitis virus (strain A59)  
N:Alternate names: protein 1b  
N:Contains: RNA-directed RNA polymerase (EC 2.7.7.48)  
C:Species: murine hepatitis virus, MHV  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 20-Jun-2000  
C:Accession: S15760; S08652  
R:Bradenbeek, P.J.; Pachuk, C.J.; Noten, A.F.H.; Charite, J.; Luytjes, W.; Weiss, S.R.;  
Nucleic Acids Res. 18, 1825-1832, 1990  
A:Title: The primary structure and expression of the second open reading frame of the po  
frameshifting mechanism.  
A:Reference number: S15760; MUID:90245573  
A:Accession: S15760  
A:Molecule type: genomic RNA  
A:Residues: 1-2733 <BRE>  
A:Cross-references: EMBL:X51939; NID:958974; PID:g1334829  
A:Superfamily: Infectious bronchitis virus RNA-directed RNA polymerase  
C:Keywords: nucleotidyltransferase; RNA biosynthesis

Query Match 42.1%; Score 45; DB 2; Length 2733;  
Best Local Similarity 52.9%; Pred. No. 1.3e+02;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 ETFOVEVPGSQHIDSOK 18  
|||||:|||||:  
Db 1198 ETQNNVPYQHIGMKR 1214

RESULT 9  
S78364  
conserved hypothetical protein 263 - Odontella sinensis chloroplast  
C:Species: chloroplast Odontella sinensis  
C:Date: 17-Feb-1998 #sequence\_revision 26-Feb-1998 #text\_change 20-Jun-2000  
C:Accession: S78364  
R:Kowallik, K.V.; Stoebe, B.; Schaffran, I.; Kroth-Pancic, P.; Freier, U.  
Plant Mol. Biol. Rep. 13, 336-342, 1995  
A:Title: The Chloroplast Genome of a chlorophyll a+c- containing Alga, Odontella sinensis  
A:Reference number: S78238  
A:Accession: S78364  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-263 <KOW>  
A:Cross-references: EMBL:Z67753; NID:g1185127; PIDN:CAA91737.1; PID:g1185254  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1995  
C:Genetics:  
A:Gene: ycf43

A:Genome: chloroplast  
C:Superfamily: conserved hypothetical protein HI0188  
C:Keywords: chloroplast

Query Match 41.1%; Score 44; DB 2; Length 263;  
Best Local Similarity 40.0%; Pred. No. 16;  
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 2 ETFOVEVPGSQHIDSOKKAI 21  
|||:|||||:  
Db 30 ETVTLELPFSEHIEELKQRL 49

RESULT 10  
T19866  
hypothetical protein C40H5.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
C:Accession: T19866  
R:White, S.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19189  
A:Accession: T19866  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-374 <WIL>  
A:Cross-references: EMBL:Z81482; PIDN:CAB03954.2; GSPDB:GN00028; CESP:C40H5.3  
A:Experimental source: clone C40H5  
C:Genetics:  
A:Gene: CESP:C40H5.3  
A:Map position: X  
A:Introns: 44/2; 95/1; 178/1; 234/3; 348/2  
C:Superfamily: Caenorhabditis elegans hypothetical protein C40H5.3

Query Match 41.1%; Score 44; DB 2; Length 374;  
Best Local Similarity 42.1%; Pred. No. 23;  
Matches 8; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHIDSOKK 19  
|||:|||||:  
Db 230 GVEFHWFPGQPHADQKE 248

RESULT 11  
T49574  
probable carnitine acetyl transferase FacC [Imported] - Neurospora crassa  
N:Alternate names: protein B208.220  
C:Species: Neurospora crassa  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C:Accession: T49574  
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu  
submitted to the Protein Sequence Database, May 2000  
A:Reference number: Z25022  
A:Accession: T49574  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-864 <SCH>  
A:Cross-references: EMBL:AL355930; GSPDB:GN00116; NCSP:B208.220  
A:Experimental source: BAC clone B208; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B208.220  
A:Map position: 6  
A:Introns: 138/1

Query Match 41.1%; Score 44; DB 2; Length 864;  
Best Local Similarity 42.1%; Pred. No. 57;  
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 ETFOVEVPGSQHIDSOKKA 20  
|||:|||||:



Db 541 QTFWADAPGDAKIDALRKA 559

**RESULT** 12

VF1HJH  
genome polyprotein 1b - murine hepatitis virus (strain JHM)  
N;Alternate names: 1b protein  
N;Contains: RNA-directed RNA polymerase (EC 2.7.7.48)  
C;Species: murine hepatitis virus, MHV  
C;Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 11-Jun-1999  
C;Accession: B36815  
R;Lee, H.J.; Shieh, C.K.; Gorbalenya, A.E.; Koonin, E.V.; La Monica, N.; Tuler, J.; Bagd  
submitted to Genbank, February 1991  
A;Description: The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding  
A;Reference number: A36815  
A;Accession: B36815  
A;Molecule type: genomic RNA  
A;Residues: 1-2731 <LEE>  
A;Cross-references: GB:M55148; NID:G331851; PID:AAA46458.1; PID:G331853  
R;Lee, H.J.; Shieh, C.K.; Gorbalenya, A.E.; Koonin, E.V.; La Monica, N.; Tuler, J.; Bagd  
Virology 180, 567-592, 1991  
A;Title: The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the  
A;Reference number: A36547; MUID:91111976  
A;Contents: annotation  
A;Note: neither nucleotide nor complete amino acid sequence is given  
C;Comment: This protein may be translated as a 1a-1b polyprotein by a ribosomal frameshift  
C;Genetics:

Query Match 41.1%; Score 44; DB 1; Length 2731;  
Best Local Similarity 69.2%; Pred. NO. 1.9e+02;  
Matches 9; Conservative 0; Mismatches 4; Indels

QY 2 ET FQVEVPGSQHI 14  
db 1198 ET FQNNVPNYQHI 12

RESULT 13

C84726  
probable receptor-like protein kinase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: C84726  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.I.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: AB4420; MUID:20083487  
A:Accession: C84726  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-641 <STO>  
A:Cross-references: GB:AE002093; NID:g4887748; PIDN:AAD32284.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g31880  
A:Map position: 2

Query Match 40.7%; Score 43.5; DB 2; Length 641;  
Best Local Similarity 50.0%; Pred. NO. 50;  
Matches 11; Conservative 3; Mismatches 7; Indels

```
QY      1 GETFQVEVPGSQ-HIDSQKKAI 21
      |||:|:|:|:|:|:|:|:|:|:|
Db     359 GEVFKAE LPGSNGKIIAVKKVI 380
```

RESULT 14

UQDOR

ubiquitin / ribosomal protein CEP52 - slime mold (Dictyostelium discoideum)  
N:Alternate names: ubiquitin fusion protein  
C:Species: Dictyostelium discoideum  
C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 22-Jun-1999  
C:Accession: S00357; A25863  
F:Muellier-Raubenberger, A.; Westphal, M.; Jaeger, E.; Noegel, A.; Gerisch, G.  
FEBS Lett. 229, 273-278, 1988  
A:Title: Complete cDNA sequence of a Dictyostelium ubiquitin with a carboxy-terminal  
A:Reference number: S00357; MUID:88152253  
A:Accession: S00357  
A:Molecule type: mRNA  
A:Residues: 1-128 <MUE>  
A:Cross-references: EMBL:X07210; NID:g7381; PIDN:CAA30183.1; PID:g7382  
A:Experimental source: strain AX2-214  
C:Gene: DUB1  
C:Superfamily: ubiquitin/ribosomal protein CEP52; ribosomal protein CEP52 homology; u  
C:Keywords: DNA binding; protein biosynthesis; protein degradation; ribosome; zinc fi  
F:1-76/Product: ubiquitin #status predicted <UBI>  
F:1-76/Domain: ubiquitin homology <UBH>  
F:77-128/Product: ribosomal protein CEP52 #status predicted <RIB>  
F:77-128/Domain: ribosomal protein CEP52 homology <CPH>  
F:95-114/Region: zinc finger CCCC motif  
F:121-128/Region: nuclear location signal

Query Match 40.2%; Score 43; DB 1; Length 128;  
Best Local Similarity 42.9%; Pred. No. 11;  
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

```
Qy 1 GETQVEVPGSQHIDSQKKAI 21
    | : | : | : | : | : |
Db 10 GKTITLEVEGSDNIENVKAKI 30
```

## RESULT 15

UQOR7  
ubiquitin / ribosomal protein S27a - slime mold (Dictyostelium discoideum)  
C:Species: Dictyostelium discoideum  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 21-Jul-2000  
C:Accession: E34080  
R:Ohmachi, T.; Giorda, R.; Shaw, D.R.; Ennis, H.L.  
Biochemistry 28, 5226-5231, 1989  
A:Title: Molecular organization of developmentally regulated Dictyostelium discoideum  
A:Reference number: A34080; MUID:89352609  
A:Accession: E34080  
A:Molecule type: mRNA  
A:Residues: 1-154 <OHM>  
A:Cross-references: GB:M23750; GB:J02858; NID:q167940; PIDN:AAA33264.1; PID:q167941  
C:Superfamily: ubiquitin / rat ribosomal protein S27a; ribosomal protein S27a homolog  
C:Keywords: protein biosynthesis; protein degradation; ribosome  
F:1-76/Product: ubiquitin #status predicted <MARI>  
F:1-76/Domain: ubiquitin homolog <UBH>  
F:77-154/Product: ribosomal protein S27a #status predicted <MAT2>  
F:102-152/Domain: ribosomal protein S27a homology <RIB>

Query Match 40.2%; Score 43; DB 1; Length 154;  
Best Local Similarity 42.9%; Pred. NO. 13;  
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

```
QY 1 GETFOVEPGSQHIDSQKKAI 21
    | : | : | | | : | : | |
DB 10 GKTITLEVEGSDNIENVKAKI 30
```

RESULT 16

ubiquitin 18 - slime mold (Dictyostelium discoideum)  
D34080  
C:Species: Dictyostelium discoideum  
C:Date: 30-Mar-1990 #sequence revision 30-Mar-1990 #

A;Note: F28A21.100

Query Match 40.2%; Score 43; DB 2; Length 368;  
Best Local Similarity 33.3%; Pred. No. 33;  
Matches 6; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 2 ETFQVEVPGSQHIDSOKK 19  
I :|:| :|:| :|:| :|:|  
Db 305 ERLPVKLPNNRHEKKEK 322

RESULT 19  
C34080  
polyubiquitin 5 (clone DCUB2) - slime mold (Dictyostellium discoideum)  
N;Alternate names: ubiquitin 2  
C;Species: Dictyostellium discoideum  
C;Date: 30-Mar-1990 #sequence\_revision 30-Mar-1990 #text\_change 11-Apr-1997  
C;Accession: C34080  
Biochemistry 28, 5226-5231, 1989  
A;Title: Molecular organization of developmentally regulated Dictyostellium discoideum  
A;Reference number: A34080; MUID:89352609  
A;Accession: C34080  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-380 <OHM>  
A;Cross-references: GB:M23754  
A;Experimental source: strain B, clone DCUB12  
C;Superfamily: polyubiquitin 5; ubiquitin homology  
C;Keywords: nucleus; polyprotein; protein degradation  
F;1-76/Product: ubiquitin #status predicted <UB1>  
F;1-76/Domain: ubiquitin homology <UBH1>  
F;77-152/Product: ubiquitin #status predicted <UB2>  
F;77-152/Domain: ubiquitin homology <UBH2>  
F;153-228/Product: ubiquitin #status predicted <UB3>  
F;153-228/Domain: ubiquitin homology <UBH3>  
F;229-304/Product: ubiquitin #status predicted <UB4>  
F;229-304/Domain: ubiquitin homology <UBH4>  
F;305-380/Product: ubiquitin #status predicted <UB5>  
F;305-380/Domain: ubiquitin homology <UBH5>

Query Match 40.2%; Score 43; DB 2; Length 380;  
Best Local Similarity 42.9%; Pred. No. 35;  
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 GETFQVEVPGSQHIDSOKKAI 21  
I:| :|:| :|:| :|:| :|:|  
Db 10 GKTITLEVEGSDNIENYKAKI 30

RESULT 20  
B34080  
polyubiquitin 5 (clone DCUB19) - slime mold (Dictyostellium discoideum)  
N;Alternate names: ubiquitin 19  
C;Species: Dictyostellium discoideum  
C;Date: 30-Mar-1990 #sequence\_revision 30-Mar-1990 #text\_change 11-Apr-1997  
C;Accession: B34080  
Biochemistry 28, 5226-5231, 1989  
A;Title: Molecular organization of developmentally regulated Dictyostellium discoideum  
A;Reference number: A34080; MUID:89352609  
A;Accession: B34080  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-380 <OHM>  
A;Cross-references: GB:M23748  
A;Experimental source: strain B, clone DCUB19  
C;Superfamily: polyubiquitin 5; ubiquitin homology  
C;Keywords: nucleus; polyprotein; protein degradation  
F;1-76/Product: ubiquitin #status predicted <UB1>  
F;1-76/Domain: ubiquitin homology <UBH1>

```

RESULT  22
A34080
polybiquitin 7 (clone DCUB14) - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 30-Mar-1990 #sequence_revision 30-Mar-1990 #text_change 11-Apr-1997
C:Accession: A34080
R:Ohmachi, T.; Giorda, R.; Shaw, H.L.
Biochemistry 28, 5226-5231, 1989
A:Title: Molecular organization of developmentally regulated Dictyostelium discoideum ub
A:Reference number: A34080; MUID:89352609
A:Accession: A34080
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-532 <OHM>
A:Cross-references: GB:M23753
A:Experimental source: strain B, clone DCUB14

```

A:Molecule type: mRNA  
A:Residues: 27-181, 'R', 'R', '183-204', 'G', '206-221', 'A', '223-380', 'D', '382-386', 'H', '388-444', 'HN', '447-  
A:Cross-references: EMBL:X08030  
A:Note: part of this sequence was confirmed by protein sequencing  
C:Superfamily: tenascin; EGF homology: fibrinogen beta/gamma homology; fibronectin type  
C:Keywords: alternative splicing; calcium binding; cell adhesion; duplication; extracellular  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-33/Domain: propeptide #status predicted <PRO>  
F:34-1810/Product: tenascin 230K #status predicted <MAT>  
F:223-249/Domain: EGF homology <EGF1>  
F:316-342/Domain: EGF homology <EGF>  
F:592-673/Domain: fibronectin type III repeat homology <FN3A>  
F:681-765/Domain: fibronectin type III repeat homology <FN3B>  
F:773-857/Domain: fibronectin type III repeat homology <FN3C>  
F:865-949/Domain: fibronectin type III repeat homology <FN3D>  
F:957-1037/Domain: fibronectin type III repeat homology <FN3E>  
F:1046-1128/Domain: fibronectin type III repeat homology <FN3F>  
F:1137-1219/Domain: fibronectin type III repeat homology <FN3G>  
F:1228-1310/Domain: fibronectin type III repeat homology <FN3H>  
F:1318-1399/Domain: fibronectin type III repeat homology <FN3I>  
F:1407-1487/Domain: fibronectin type III repeat homology <FN3J>  
F:1495-1575/Domain: fibronectin type III repeat homology <FN3K>  
F:1590-1798/Domain: fibrinogen beta/gamma homology <FBG>  
F:1734-1747/Domain: calcium binding #status predicted <CAB>

Query Match 40.2%; Score 43; DB 1; Length 1810;  
Best Local Similarity 58.3%; Pred. No. 1.8e+02;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 ETFQVEVPGSQH 13  
||| : ||| ||  
DB 1089 ETWNTVPGQH 1100

## RESULT 24

E82963  
probable peptide chain release factor PA5470 [imported] - Pseudomonas aeruginosa (strain  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: E82963  
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337  
A:Accession: E82963  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-204 <STO>  
A:Cross-references: GB:AE004959; GB:AE004091; NID:g9951791; PIDN:RAG08855.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA5470

Query Match 39.3%; Score 42; DB 2; Length 204;  
Best Local Similarity 35.0%; Pred. No. 26;  
Matches 7; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 2 ETFQVEVPGSQHDSQKAI 21  
||| : ||| ||| :  
DB 112 ETLRSSGPGGHVNTDSAV 131

## RESULT 25

JC5694  
stress-activated protein kinase (EC 2.7.-.-) JNKB - common carp  
N:Alternate names: stress-activated protein kinase b  
C:Species: Cyprinus carpio (common carp)  
C:Date: 20-Nov-1997 #sequence\_revision 20-Nov-1997 #text\_change 24-Sep-1999  
C:Accession: JC5694

R:Hashimoto, H.; Matsuo, Y.; Yokoyama, Y.; Toyohara, H.; Sakaguchi, M.  
J. Biochem. 122, 381-386, 1997  
A:Title: Structure and expression of carp mitogen-activated protein kinases homologou  
A:Reference number: JC5693; MUID:97456373  
A:Accession: JC5694  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-427 <HAS>  
A:Cross-references: DDBJ:AB001744; NID:g2467307; PIDN:BAA22598.1; PID:d1023466; PID:g  
C:Comment: This enzyme is a mitogen-activated protein kinase, and plays a role in egg  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom  
C:Keywords: phosphotransferase  
F:24-278/Domain: protein kinase homology <KIN>

Query Match 39.3%; Score 42; DB 2; Length 427;  
Best Local Similarity 61.5%; Pred. No. 57;  
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 9 PGSDHDSQKAI 21  
||| ||| |||  
DB 226 PGSDHIDQWNV 238

Search completed: July 16, 2001, 16:37:02  
Job time: 211 sec



***This Page Blank (uspto)***

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 16:44:04 ; Search time 19.88 Seconds  
(without alignments)  
36.185 Million cell updates/sec

Title: US-09-786-648-5  
Perfect score: 107  
Sequence: 1 GETFQVEVPGSQHDSQKAI 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	100.0	124	1 ELBP_ECOLI	P32890 escherichia
2	101	94.4	124	1 ELBH_ECOLI	P13811 escherichia
3	95	88.8	124	1 CHTB_VIBCH	P01556 vibrio chol
4	45	42.1	2733	1 RRPE_CVMA5	P16342 murine coro
5	44	41.1	263	1 YC43_ODOSI	P49538 odontella s
6	44	41.1	467	1 IRF6_HUMAN	O14896 homo sapien
7	44	41.1	2731	1 RRPE_CVMJH	P29982 murine coro
8	43	40.2	76	1 UBIQ_DICDI	P08618 dictyosteli
9	43	40.2	1808	1 TENA_CHICK	P10039 gallus gall
10	42	39.3	855	1 STI4_HUMAN	Q9Y3Y6 homo sapien
11	41.5	38.8	376	1 AROC_YEAST	P28777 saccharomyc
12	41.5	38.8	500	1 GABT_HUMAN	P80404 homo sapien
13	41	38.3	141	1 RFH_ECOLI	P28369 escherichia
14	41	38.3	313	1 YP68_CAEEL	Q09217 caenorhabdi
15	41	38.3	385	1 TGT_ZYMO	P28720 zymomonas m
16	41	38.3	386	1 YAAH_BACSU	P37535 bacillus su
17	41	38.3	683	1 RPC_BPPHC	P08979 bacterioph
18	41	38.3	1702	1 DPOL_THELI	P30317 thermococcu
19	41	38.3	1829	1 DPOL_THEST	O33845 thermococcu
20	40.5	37.9	862	1 PMS2_HUMAN	P54278 homo sapien
21	40	37.4	76	1 UBIQ_NEUCR	P13117 neurospora
22	40	37.4	148	1 IF5A_AERPE	Q9Y453 aeropyrum p
23	40	37.4	288	1 HS74_CAEEL	P20163 caenorhabdi
24	40	37.4	355	1 YRY1_CAEEL	Q10005 caenorhabdi
25	40	37.4	395	1 ADH3_ENTHI	Q24857 entamoeba h
26	40	37.4	523	1 DBP3_YEAST	P20447 saccharomyc
27	40	37.4	773	1 DPOL_THEGO	P56889 thermococcu
28	40	37.4	775	1 DPOL_THESE9	Q56366 thermococcu
29	40	37.4	1048	1 CPXB_BACME	P14779 bacillus me
30	40	37.4	1523	1 DPOL_THEFM	P74918 thermococcu
31	39.5	36.9	1090	1 PULA_KLEPN	P07206 klebsiella
32	39	36.4	76	1 UBIQ_YEAST	P04838 saccharomyc
33	39	36.4	271	1 CEAM_ECOLI	P05820 escherichia

34 39 36.4 320 1 Y054\_MYCPN P75049 mycoplasma  
35 39 36.4 445 1 RF1M\_HUMAN O75570 homo sapien  
36 39 36.4 456 1 SR54\_THBAC O9hk00 thermoplas  
37 39 36.4 459 1 IL7R\_MOUSE P16872 mus musculu  
38 39 36.4 491 1 CD5\_RAT P51882 rattus norv  
39 39 36.4 560 1 INR1\_SHEEP O28599 ovis aries  
40 39 36.4 619 1 BCHD\_CHLVI O50313 chlorobium  
41 39 36.4 661 1 HS7C\_CAEEL P27430 caenorhabdi  
42 39 36.4 684 1 CC4\_CANAL P53699 candida alb  
43 39 36.4 704 1 PCCA\_RAT P14882 rattus norv  
44 39 36.4 775 1 TH11\_SCHPO P36598 schizosacch  
45 39 36.4 842 1 DPOL\_HPBVM P31870 hepatitis b

## ALIGNMENTS

RESULT 1  
ELBP\_ECOLI STANDARD; PRT; 124 AA.  
AC P32890; P13768; P01557;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE HEAT-LABILE ENTEROTOXIN B CHAIN PRECURSOR (LT-B; PORCINE) (LTP-B).  
GN ELTB OR LTPB.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ISOLATE P307;  
RX MEDLINE=81074965; PubMed=7003397;  
RA Dallas W.S., Falkow S.;  
RT "Amino acid sequence homology between cholera toxin and Escherichia coli heat-labile toxin.";  
RT Nature 288:499-501(1980).  
RN [2]  
RP REVISIONS TO 28 AND 64.  
RC STRAIN=ISOLATE P307;  
RX MEDLINE=85156481; PubMed=3884513;  
RA Leong J., Vinal A.C., Dallas W.S.;  
RT "Nucleotide sequence comparison between heat-labile toxin B-subunit cistrons from Escherichia coli of human and porcine origin.";  
RL Infect. Immun. 48:73-77(1985).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ISOLATE PCG86;  
RX MEDLINE=87137303; PubMed=3546273;  
RA Yamamoto T., Gojobori T., Yokota T.;  
RT "Evolutionary origin of pathogenic determinants in enterotoxigenic Escherichia coli and Vibrio cholerae O1.";  
RL J. Bacteriol. 169:1352-1357(1987).  
RN [4]  
RP SEQUENCE OF 1-22 FROM N.A.  
RC MEDLINE=87280041; PubMed=3301830;  
RA Ibrahim I., Gentz R.;  
RT "A functional interaction between the signal peptide and the translation apparatus is detected by the use of a single point mutation which blocks translocation across mammalian endoplasmic reticulum.";  
RL J. Biol. Chem. 262:10189-10194(1987).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).  
RX MEDLINE=93240541; PubMed=8478941;  
RA Sixma T.K., van Zanten B.A.M., Dauter Z., Hol W.G.J.;  
RT "Refined structure of Escherichia coli heat-labile enterotoxin, a close relative of cholera toxin.";  
RL J. Mol. Biol. 230:890-918(1993).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
RX MEDLINE=91238966; PubMed=2034287;

RA Sixma T.K., Pronk S.E., Kalk K.H., Wartna E.S., van Zanten B.A.M.,  
RT "Crystal structure of a cholera toxin-related heat-labile enterotoxin  
from E. coli.",  
RN Nature 351:371-377(1991).  
RX DISCUSSION OF SEQUENCE.  
RA MEDLINE-95349400; PubMed-7623669;  
RA Domenighini M., Pizzo M., Jobling M.G., Holmes R.K., Rappuoli R.;  
RT "Identification of errors among database sequence entries and  
comparison of correct amino acid sequences for the heat-labile  
enterotoxins of Escherichia coli and Vibrio cholerae.",  
RL Mol. Microbiol. 15:1165-1167(1995).  
CC -!- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY  
CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.  
CC -!- SUBUNIT: HETEROXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M17873; AAA98065.1; -;  
DR EMBL; M15363; AAA24792.1; -;  
DR EMBL; M17101; AAA23973.1; -;  
DR PIR; A01820; QLECB.  
DR PIR; B26946; QLECEB.  
DR PDB; 1LTA; 31-JAN-94.  
DR PDB; 1LTB; 31-JAN-94.  
DR PDB; 1LTG; 15-SEP-95.  
DR PDB; 1LTI; 17-AUG-96.  
DR PDB; 1LTS; 31-JAN-94.  
DR PDB; 1LTT; 31-JAN-94.  
DR PDB; 1LTU; 07-JUL-97.  
DR PDB; 1LT4; 16-JUN-97.  
DR PDB; 1LT5; 03-DEC-97.  
DR PDB; 1LT6; 03-DEC-97.  
DR PDB; 1LTL; 20-APR-95.  
DR InterPro; IPR001835; -;  
DR Pfam; PF01376; Enterotoxin\_B; 1.  
DR PRINTS; PR00772; ENTEROTOXINB.  
KW Enterotoxin; Signal; 3D-structure.  
FT SIGNAL 1 21  
FT CHAIN 22 124 HEAT-LABILE ENTEROTOXIN B CHAIN.  
FT DISULFID 30 107  
FT HELIX 26 30  
FT TURN 31 32  
FT STRAND 36 43  
FT STRAND 47 51  
FT TURN 54 55  
FT STRAND 58 62  
FT TURN 64 65  
FT STRAND 68 71  
FT TURN 76 77  
FT HELIX 80 98  
FT TURN 99 100  
FT STRAND 103 109  
FT STRAND 115 123  
SQ SEQUENCE 124 AA; 14133 MW; 6DB7DE58395EA70D CRC64;

Query Match 100.0%; Score 107; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 3.2e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GETFOVEVPGSOHIDSQKKAI 21  
DB 66 GETFOVEVPGSOHIDSQKKAI 86

RESULT 2  
ELBH\_ECOLI STANDARD; PRT; 124 AA.  
AC PI3811;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE HEAT-LABILE ENTEROTOXIN B CHAIN PRECURSOR (LT-B, HUMAN) (LTH-B).  
GN ELTB OR LTPB.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ISOLATE H74-114;  
RX MEDLINE-85156481; PubMed-3884513;  
RA Leong J., Vinal A.C., Dallas W.S.;  
RT "Nucleotide sequence comparison between heat-labile toxin B-subunit  
cistrons from Escherichia coli of human and porcine origin.",  
RL Infect. Immun. 48:73-77(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ISOLATE H10407;  
RX MEDLINE-83114628; PubMed-6759877;  
RA Yamamoto T., Tamura T.A., Yokota T., Takano T.;  
RT "Overlapping genes in the heat-labile enterotoxin operon originating  
from Escherichia coli human strain.",  
RL Mol. Gen. Genet. 188:356-359(1982).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ISOLATE H10407;  
RX MEDLINE-93252225; PubMed-8486242;  
RA Inoue T., Tsuji T., Koto M., Imamura S., Miyama A.;  
RT "Amino acid sequence of heat-labile enterotoxin from chicken  
enterotoxigenic Escherichia coli is identical to that of human strain  
H 10407.",  
RL FEMS Microbiol. Lett. 108:157-161(1993).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ETEC LT 87;  
RA Germani Y., Desperrier J.M.;  
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP DISCUSSION OF SEQUENCE.  
RX MEDLINE-95349400; PubMed-7623669;  
RA Domenighini M., Pizzo M., Jobling M.G., Holmes R.K., Rappuoli R.;  
RT "Identification of errors among database sequence entries and  
comparison of correct amino acid sequences for the heat-labile  
enterotoxins of Escherichia coli and Vibrio cholerae.",  
RL Mol. Microbiol. 15:1165-1167(1995).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (3.04 ANGSTROMS).  
RX MEDLINE-99185101; PubMed-10085117;  
RA Matkovic-Calogovic D., Loregian A., D'Acunto M.R., Battistutta R.,  
RA Tossi A., Palu G., Zanotti G.;  
RT "Crystal structure of the B subunit of escherichia coli heat-labile  
enterotoxin carrying peptides with anti-herpes simplex virus type 1  
activity.",  
RL J. Biol. Chem. 274:8764-8769(1999).  
CC -!- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY  
CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.  
CC -!- SUBUNIT: HETEROXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M17874; AAA98064.1; -;



[6]  
SEQUENCE FROM N.A.  
STRAIN=EL TOR N16961 / SEROTYPE O1;  
MEDLINE=20406833; PubMed=10952301;  
Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
Fraser C.M.;  
"DNA sequence of both chromosomes of the cholera pathogen *Vibrio*  
*cholerae*";  
Nature 406:477-483(2000).  
[7]  
SEQUENCE OF 22-124.  
MEDLINE=78005537; PubMed=903363;  
Kurosky A., Markel D.E., Peterson J.W.;  
"Covalent structure of the beta chain of cholera enterotoxin";  
J. Biol. Chem. 252:7257-7264(1977).  
[8]  
SEQUENCE OF 22-124.  
MEDLINE=78005536; PubMed=903362;  
Lai C.-Y.;  
"Determination of the primary structure of cholera toxin B subunit.";  
J. Biol. Chem. 252:7249-7256(1977).  
[9]  
X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).  
MEDLINE=94272319; PubMed=8003954;  
Merritt E.A., Sarfaty S., van den Akker F., L'Hoir C., Martial J.A.,  
Hol W.G.J.;  
"Crystal structure of cholera toxin B-pentamer bound to receptor GM1  
pentasaccharide";  
Protein Sci. 3:166-175(1994).  
[10]  
X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
MEDLINE=95387394; PubMed=7658472;  
Zhang R.G., Westbrook M.L., Westbrook E.M., Scott D.L., Otwinowski Z.,  
Maulik P.R., Reed R.A., Shipley G.G.;  
"The 2.4 Å crystal structure of cholera toxin B subunit pentamer:  
choleragenoid";  
J. Mol. Biol. 251:550-562(1995).  
[11]  
X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
STRAIN=OGAWA 41 / CLASSICAL BIOTYPE;  
MEDLINE=97376625; PubMed=9232653;  
Merritt E.A., Sarfaty S., Jobling M.G., Chang T., Holmes R.K.,  
Hirst T.R., Hol W.G.J.;  
"Structural studies of receptor binding by cholera toxin mutants";  
Protein Sci. 6:1516-1528(1997).  
CC -!- FUNCTION: THE BETA CHAIN AGGREGATE (B SUBUNIT) IS INVOLVED IN  
BINDING TO CELL MEMBRANES.  
CC -!- SUBUNIT: CONTAINS 3 KINDS OF CHAINS. AN ALPHA AND A GAMMA CHAIN  
(FROM THE SAME PRECURSOR MOLECULE), LINKED BY AN INTERCHAIN  
DISULFIDE BOND, ASSOCIATE NONCOVALENTLY WITH AN AGGREGATE OF 4 TO  
6 BETA CHAINS.

[illegible]

```
DR PIR; A05130; A05130.
DR PIR; S14624; S14624.
DR PDB; 2CHB; 03-DEC-97.
DR PDB; 3CHB; 12-AUG-98.
DR PDB; 1CHP; 08-MAR-96.
DR PDB; 1CHQ; 08-MAR-96.
DR PDB; 1FGB; 23-DEC-96.
DR PDB; 1XPB; 01-APR-97.
DR PDB; 1XTC; 01-AUG-96.
DR PDB; 1CT1; 15-OCT-97.
DR TIGR; VC1456; -.
DR InterPro; IPR001835; -.
DR Pfam; PF01376; Enterotoxin_B; 1.
DR PRINTS; PR00772; ENTEROTOXINB.
KW Membrane; Enterotoxin; Signal; 3D-structure.
FT SIGNAL 1 21
FT CHAIN 22 124
FT DISULFID 30 107
FT CONFLICT 33 33
FT CONFLICT 39 39
FT CONFLICT 43 43
FT CONFLICT 68 68
FT CONFLICT 70 70
FT CONFLICT 75 75
FT CONFLICT 91 91
FT CONFLICT 95 95
FT HELIX 26 30
FT TURN 31 32
FT TURN 34 35
FT STRAND 36 44
FT STRAND 47 51
FT TURN 54 55
FT STRAND 58 62
FT TURN 64 65
FT STRAND 68 71
FT TURN 76 77
FT TURN 80 99
FT TURN 100 100
FT STRAND 102 109
FT STRAND 115 123
SQ SEQUENCE 124 AA; 13957 MW; 9AA393E3EA8E3EBF CRC64;

Query Match 88.8%; Score 95; DB 1; Length 124;
Best Local Similarity 90.5%; Pred. No. 2.6e-08;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GETFQVEVPGSQHDSQKAI 21
   | ||||| ||||| |||||
DB 66 GAIFQVEVPGSQHDSQKAI 86

RESULT 4
ID RRPB_CVMA5 STANDARD; PRT; 2733 AA.
AC P16342;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (ORF1B).
GN F2.
OS Murine coronavirus MHV (strain A59).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=11142;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90245573; PubMed=2159623;
RA Bredenbeek P.J., Pachuk C.J., Noten A.F.H., Charite J., Luytjes W.,
RA Weiss S.R., Spaan W.J.M.;
RT "The primary structure and expression of the second open reading
RT frame of the polymerase gene of the coronavirus MHV-A59; a highly
RT conserved polymerase is expressed by an efficient ribosomal
RT frameshifting mechanism.";
```

DR Pfam: PF00902; UPF0032; 1.  
DR PROSITE: PS01218; TATC; 1.  
KW Chloroplast; Hypothetical protein; Transmembrane.

FT TRANSMEM 53 73 POTENTIAL.  
FT TRANSMEM 103 123 POTENTIAL.  
FT TRANSMEM 130 150 POTENTIAL.  
FT TRANSMEM 153 173 POTENTIAL.  
FT TRANSMEM 181 201 POTENTIAL.  
FT TRANSMEM 213 233 POTENTIAL.  
FT TRANSMEM 241 261 POTENTIAL.  
SQ SEQUENCE 263 AA; 30080 MW; 25083364F2A89E88 CRC64;

Query Match 41.1%; Score 44; DB 1; Length 263;  
Best Local Similarity 40.0%; Pred. No. 7.8;  
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 2 ETEFVEVPGSQHIDSOKKAI 21  
II :I:I I:I I: :  
Db 30 ETVTLELPFSEHIEELKQRL 49

RESULT 6  
IRF6\_HUMAN  
ID IRF6\_HUMAN STANDARD; PRT; 467 AA.  
AC Q14896;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE INTERFERON REGULATORY FACTOR 6 (IRF-6).  
GN IRF6.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]

RP SEQUENCE FROM N.A.  
RA Grossman A., Mittrucker H.W., Antonio L., Ozato K., Mak T.W.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]

RP SEQUENCE FROM N.A.  
RA Grafham D.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE IRF FAMILY.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: AF027292; AAB84111.1; -;  
DR EMBL: AL022398; CAA18545.1; -;  
DR HSP: P23906; IIRG.  
DR InterPro: IPR001346; -;  
DR Pfam: PF00605; IRF; 1.  
DR PRINTS: PR00267; INTERREGFCT.  
DR PROSITE: PS00601; IRF; 1.  
KW Transcription regulation; DNA-binding; Nuclear protein.  
FT DNA\_BIND 9 111 TRYPTOPHAN PENTAD REPEAT.  
SQ SEQUENCE 467 AA; 53129 MW; 7E28F5E0F5BA4053 CRC64;

Query Match 41.1%; Score 44; DB 1; Length 467;  
Best Local Similarity 41.2%; Pred. No. 14;  
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 5 QVEVPGSQHIDSOKKAI 21  
II :I:I I:I I: :  
Db 273 QVKEPGPEHTNEKQL 289

RESULT 7  
RRPB\_CVMJH

ID RRPB\_CVMJH STANDARD; PRT; 2731 AA.  
AC P29982;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (ORF1B).  
OS Murine coronavirus MHV (strain JHM).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
OC Coronaviridae; Coronavirus.  
OX NCBI\_TaxID=11144;  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE=9111176; PubMed=1846489;  
RA Lee H.-J., Shieh C.-K., Gorbalenya A.E., Koonin E.V., la Monica N.,  
RA Tuler J., Bagdzhardzhan A., Lai M.M.C.;  
RT "The complete sequence (22 kilobases) of murine coronavirus gene 1  
RT encoding the putative proteases and RNA polymerase.";  
RL Virology 180:567-582(1991).

CC -!- FUNCTION: THE RNA DEPENDENT RNA POLYMERASE OF CORONAVIRUSES IS  
CC A MULTIFUNCTIONAL PROTEIN: IT CONTAINS THE ACTIVITIES NECESSARY  
CC FOR THE TRANSCRIPTION OF NEGATIVE STRANDED RNA, LEADER RNA,  
CC SUBGENOMIC MRNAS AND PROGENY VIRION RNA.  
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -> N PYROPHOSPHATE  
CC + RNA(N).  
CC -!- MISCELLANEOUS: THIS PROTEIN MAY BE TRANSLATED AS A 1A-1B  
CC POLYPROTEIN BY A RIBOSOMAL FRAMESHIFTING MECHANISM.  
CC -!- SIMILARITY: TO THE AVIAN CORONAVIRUS IBV RNA POLYMERASE.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: M55148; AAA46458.1; -;  
DR PIR: B36815; VFIHJH.  
KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.  
FT DOMAIN 622 824 POLYMERASE.  
FT DOMAIN 944 1014 CYS/HIS-RICH.  
FT DOMAIN 1218 1506 HELICASE.  
FT NP\_BIND 1220 1227 ATP (BY SIMILARITY).  
SQ SEQUENCE 2731 AA; 308834 MW; 73B3C4025244CEE5 CRC64;

Query Match 41.1%; Score 44; DB 1; Length 2731;  
Best Local Similarity 69.2%; Pred. No. 95;  
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ETEFVEVPGSQHI 14  
IIII I I I I I  
Db 1198 ETEFQNNVPYQHI 1210

RESULT 8  
UBIQ\_DICDI

ID UBIQ\_DICDI STANDARD; PRT; 76 AA.  
AC P08618;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 01-NOV-1991 (Rel. 20, Last annotation update)  
DE UBIQUITIN.

OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88152253; PubMed=2831095;



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; M33121; AAA49086.1; -;  
DR EMBL; X08031; CAB40811.1; -;  
DR EMBL; X08030; CAA30824.1; ALT\_TERM.  
DR EMBL; J03641; AAA48748.1; ALT\_SEQ.  
DR EMBL; M20816; AAA48749.1; ALT\_SEQ.  
DR PIR; A30903; A30903.  
DR PIR; A33930; A31930.  
DR PIR; A33379; A33379.  
DR PIR; B33379; B33379.  
DR PIR; C33379; C33379.  
DR PIR; S01292; S01292.  
DR HSP; P24821; ITEN.  
DR InterPro; IPR000561; -;  
DR InterPro; IPR001777; -;  
DR InterPro; IPR002181; -;  
DR Pfam; PF00008; EGF\_13.  
DR Pfam; PF00147; fibrinogen\_C; 1.  
DR Pfam; PF00041; fn3; 11.  
DR PROSITE; PS00022; EGF\_1; 14.  
DR PROSITE; PS01186; EGF\_2; 14.  
KW Glycoprotein; Cell adhesion; Repeat; EGF-like domain; Coiled coil;  
KW Extracellular matrix; Signal; Alternative splicing.  
FT SIGNAL 1 22  
FT PROPEP 23 33  
FT CHAIN 34 1808  
FT DOMAIN 118 142 TENASCIN.  
FT DOMAIN 176 188 COILED COIL (POTENTIAL).  
FT DOMAIN 188 219 EGF-LIKE 1 (INCOMPLETE).  
FT DOMAIN 219 250 EGF-LIKE 2.  
FT DOMAIN 250 281 EGF-LIKE 3.  
FT DOMAIN 281 312 EGF-LIKE 4.  
FT DOMAIN 312 343 EGF-LIKE 5.  
FT DOMAIN 343 374 EGF-LIKE 6.  
FT DOMAIN 374 405 EGF-LIKE 7.  
FT DOMAIN 405 436 EGF-LIKE 8.  
FT DOMAIN 436 467 EGF-LIKE 9.  
FT DOMAIN 467 498 EGF-LIKE 10.  
FT DOMAIN 498 529 EGF-LIKE 11.  
FT DOMAIN 529 560 EGF-LIKE 12.  
FT DOMAIN 560 591 EGF-LIKE 13.  
FT DOMAIN 591 680 EGF-LIKE 14.  
FT DOMAIN 681 771 FIBRONECTIN TYPE-III 1.  
FT DOMAIN 772 862 FIBRONECTIN TYPE-III 2.  
FT DOMAIN 863 954 FIBRONECTIN TYPE-III 3.  
FT DOMAIN 955 1042 FIBRONECTIN TYPE-III 4.  
FT DOMAIN 1043 1133 FIBRONECTIN TYPE-III 5.  
FT DOMAIN 1134 1224 FIBRONECTIN TYPE-III 6.  
FT DOMAIN 1225 1315 FIBRONECTIN TYPE-III 7.  
FT DOMAIN 1316 1404 FIBRONECTIN TYPE-III 8.  
FT DOMAIN 1405 1492 FIBRONECTIN TYPE-III 9.  
FT DOMAIN 1493 1580 FIBRONECTIN TYPE-III 10.  
FT DOMAIN 1589 1808 FIBRONECTIN TYPE-III 11.  
FT DISULFID 64 64 FIBRINOGEN BETA/GAMMA.  
FT DISULFID 192 202 INTERCHAIN (POTENTIAL).  
FT DISULFID 196 207 BY SIMILARITY.  
FT DISULFID 209 218 BY SIMILARITY.  
FT DISULFID 223 233 BY SIMILARITY.  
FT DISULFID 227 238 BY SIMILARITY.  
FT DISULFID 240 249 BY SIMILARITY.  
FT DISULFID 254 264 BY SIMILARITY.  
FT DISULFID 258 269 BY SIMILARITY.  
FT DISULFID 271 280 BY SIMILARITY.  
FT DISULFID 285 295 BY SIMILARITY.  
FT DISULFID 289 300 BY SIMILARITY.  
FT DISULFID 302 311 BY SIMILARITY.

FT DISULFID 316 326 BY SIMILARITY.  
FT DISULFID 320 331 BY SIMILARITY.  
FT DISULFID 333 342 BY SIMILARITY.  
FT DISULFID 347 357 BY SIMILARITY.  
FT DISULFID 351 362 BY SIMILARITY.  
FT DISULFID 364 373 BY SIMILARITY.  
FT DISULFID 378 388 BY SIMILARITY.  
FT DISULFID 382 393 BY SIMILARITY.  
FT DISULFID 395 404 BY SIMILARITY.  
FT DISULFID 409 419 BY SIMILARITY.  
FT DISULFID 413 424 BY SIMILARITY.  
FT DISULFID 426 435 BY SIMILARITY.  
FT DISULFID 440 450 BY SIMILARITY.  
FT DISULFID 444 455 BY SIMILARITY.  
FT DISULFID 457 466 BY SIMILARITY.  
FT DISULFID 471 481 BY SIMILARITY.  
FT DISULFID 475 486 BY SIMILARITY.  
FT DISULFID 488 497 BY SIMILARITY.  
FT DISULFID 502 512 BY SIMILARITY.  
FT DISULFID 506 517 BY SIMILARITY.  
FT DISULFID 519 528 BY SIMILARITY.  
FT DISULFID 533 543 BY SIMILARITY.  
FT DISULFID 537 548 BY SIMILARITY.  
FT DISULFID 550 559 BY SIMILARITY.  
FT DISULFID 564 574 BY SIMILARITY.  
FT DISULFID 568 579 BY SIMILARITY.  
FT DISULFID 581 590 BY SIMILARITY.  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 168 168 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 186 186 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 328 328 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 603 603 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 643 643 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 751 751 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 759 759 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1050 1050 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1090 1090 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1101 1101 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1112 1112 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1153 1153 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1183 1183 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1416 1416 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1736 1736 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1769 1769 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT VARSPLIC 1043 1224 MISSING (IN 200 KDA ISOFORM).  
FT VARSPLIC 1043 1315 MISSING (IN 190 KDA ISOFORM).  
FT CONFLICT 182 182 W -> R (IN REF. 2).  
FT CONFLICT 563 571 SCPNDCNNV -> PAPMTATTW (IN REF. 3).  
FT CONFLICT 598 598 E -> G (IN REF. 3).  
FT CONFLICT 838 838 T -> Y (IN REF. 3).  
FT CONFLICT 886 886 N -> F (IN REF. 3).  
SQ SEQUENCE 1808 AA; 198858 MW; B924A06CF9EFD6DE CRC64;  
Query Match 40.2%; Score 43; DB 1; Length 1808;  
Best Local Similarity 58.3%; Pred. No. 88;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 2 ETFQVEVPGSQH 13  
||: : ||| ||  
Db 1087 ETWNITVPGSQH 1098  
RESULT 10  
ST14\_HUMAN STANDARD; PRT; 855 AA.  
ID ST14\_HUMAN  
AC Q9Y5Y6;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE SUPPRESSOR OF TUMORIGENICITY 14 (EC 3.4.21.-) (MATRIPITASE) (MEMBRANE-  
DE TYPE SERINE PROTEASE 1) (MT-SP1).  
GN ST14 OR PRSS14 OR SNCL19.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99303581; PubMed=10373424;  
 RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;  
 RT "Molecular cloning of cDNA for matrilysin, a matrix-degrading serine  
 protease with trypsin-like activity.";  
 RL J. Biol. Chem. 274:18231-18236(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Takeuchi T., Shuman M.A., Craik C.S.;  
 RT "Reverse biochemistry: Use of macromolecular protease inhibitors to  
 dissect complex biological processes and identify a membrane-type  
 serine protease in epithelial cancer and normal tissue.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).  
 RN [3]  
 RP CHARACTERIZATION.  
 RC TISSUE=Milk;  
 RX PubMed=10373425;  
 RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;  
 RT "Purification and characterization of a complex containing matrilysin  
 and a kunitz-type serine protease inhibitor from human milk.";  
 RL J. Biol. Chem. 274:18237-18242(1999).  
 CC -!- FUNCTION: DEGRADES EXTRACELLULAR MATRIX. PROPOSED TO PLAY A ROLE  
 CC IN BREAST CANCER INVASION AND METASTASIS. EXHIBITS TRYPSIN-LIKE  
 CC ACTIVITY AS DEFINED BY CLEAVAGE OF SYNTHETIC SUBSTRATES WITH ARG  
 CC OR LYS AS THE P1 SITE.  
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).  
 CC -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF118224; AAD42765.2; -;  
 DR EMBL; AF133086; AAF00109.1; -;  
 DR HSSP; P00763; IDPO.  
 DR MEROPS; S01.302; -;  
 DR InterPro; IPR000859; -;  
 DR InterPro; IPR001254; -;  
 DR InterPro; IPR001314; -;  
 DR InterPro; IPR002172; -;  
 DR Pfam; PF00057; ldl\_recept\_a; 4.  
 DR Pfam; PF00089; trypsin; 1.  
 DR Pfam; PF00431; CUB; 2.  
 DR PRINTS; PR00261; LDLRECEPTOR.  
 DR PROSITE; PR00722; CHYMOTRYPSIN.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR PROSITE; PS01180; CUB; 2.  
 DR PROSITE; PS01209; LDLRA\_1; 2.  
 DR PROSITE; PS00068; LDLRA\_2; 4.  
 KW Signal-anchor; Glycoprotein; Hydrolase; Serine protease;  
 KW Transmembrane; Repeat.  
 FT DOMAIN 1 55 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 56 76 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT EXTRACELLULAR (POTENTIAL).  
 FT CUB 1.  
 FT CUB 2.  
 FT LDL-RECEPTOR CLASS A 1.  
 FT LDL-RECEPTOR CLASS A 2.  
 FT LDL-RECEPTOR CLASS A 3.

FT DOMAIN 566 603 LDL-RECEPTOR CLASS A 4.  
 FT ACT\_SITE 614 851 CATALYTIC.  
 FT ACT\_SITE 656 656 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 711 711 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 805 805 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 772 772 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 855 AA; 94769 MW; 26143132C01F99C9 CRC64;  
 Query Match 39.3%; Score 42; DB 1; Length 855;  
 Best Local Similarity 50.0%; Pred. No. 57;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
 QY 3 TFOVEVPGSQHI 14  
 DB 367 TWNIEVPNNQHV 378  
 RESULT 11  
 AROC\_YEAST  
 ID AROC\_YEAST STANDARD; PRT; 376 AA.  
 AC P28777;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE CHORISMATE SYNTHASE (EC 4.6.1.4) (5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE  
 DE PHOSPHOLYASE).  
 GN ARO2 OR YGL148W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RX MEDLINE=92114793; PubMed=1837329;  
 RA Jones D.G.B., Reusser U., Braus G.H.;  
 RT "Molecular cloning, characterization and analysis of the regulation  
 RT of the ARO2 gene, encoding chorismate synthase, of Saccharomyces  
 RT cerevisiae.";  
 RL Mol. Microbiol. 5:2143-2152(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / FY1769;  
 RX MEDLINE=97197983; PubMed=9046099;  
 RA Voet M., Defoor E., Verhasselt P., Riles L., Robben J., Volckaert G.;  
 RT "The sequence of a nearly unclonable 22.8 kb segment on the left arm  
 RT chromosome VII from Saccharomyces cerevisiae reveals ARO2, RPL9A,  
 RT TIF1, MRF1 genes and six new open reading frames.";  
 RL Yeast 13:177-182(1997).  
 CC -!- CATALYTIC ACTIVITY: 5-O-(1-CARBOXYVINYL)-3-PHOSPHOSHIKIMATE -  
 CC CHORISMATE + ORTHOPHOSPHATE.  
 CC -!- COFACTOR: REDUCED FLAVIN.  
 CC -!- PATHWAY: SEVENTH STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN  
 CC THE BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).  
 CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -!- INDUCTION: BY AMINO ACID STARVATION.  
 CC -!- SIMILARITY: BELONGS TO THE CHORISMATE SYNTHASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X60190; CAA42745.1; -;  
 DR EMBL; X99960; CAA68214.1; -;  
 DR EMBL; Z72670; CAA96860.1; -;

```

DR PIR: S17246; S17246.
DR SGD: S0003116; ARO2.
DR InterPro: IPR000453; -.
DR Pfam: PF01264; Chorismate_synt; 1.
DR PROSITE: PS00787; CHORISMATE SYNTHASE.1; 1.
DR PROSITE: PS00788; CHORISMATE SYNTHASE.2; 1.
DR PROSITE: PS00789; CHORISMATE SYNTHASE.3; 1.
KW Lyase; Aromatic amino acid biosynthesis.
SQ SEQUENCE 376 AA; 40838 MW; AF3AF65605B91E8E CRC64;

Query Match 38.8%; Score 41.5; DB 1; Length 376;
Best Local Similarity 62.5%; Pred. No. 29;
Matches 10; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 1 GETFQ-VEVPGSHID 15
Db 258 GSGFGVSGVPGSKND 273

RESULT 12
GABT_HUMAN STANDARD; PRT; 500 AA.
AC P80404;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 4-AMINO-BUTYRATE AMINOTRANSFERASE, MITOCHONDRIAL PRECURSOR
DE (EC 2.6.1.19) (GAMMA-AMINO-N-BUTYRATE TRANSAMINASE) (GABA
DE TRANSAMINASE) (GABA AMINOTRANSFERASE) (GABA-AT).
GN ABAT OR GABAT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95237607; PubMed=7721088;
RA Osei Y.D., Churchich J.E.;
RT "Screening and sequence determination of a cDNA encoding the human
RT brain 4-aminobutyrate aminotransferase.";
RL Gene 155:185-187(1995).
RN [2]
RP SEQUENCE OF 368-465 FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=95154329; PubMed=7851425;
RA de Blase D., Barra D., Simmaco M., John R.A., Bossa F.;
RT "Primary structure and tissue distribution of human 4-aminobutyrate
RT aminotransferase.";
RL Eur. J. Biochem. 227:476-480(1995).
RN [3]
RP VARIANT LYS-220.
RX MEDLINE=99336116; PubMed=10407778;
RA Medina-Kauwe L.K., Tobin A.J., De Meirleir L., Jaeken J., Jakobs C.,
RA Nyhan W.L., Gibson K.M.;
RT "4-aminobutyrate aminotransferase (GABA-transaminase) deficiency.";
RL J. Inher. Metab. Dis. 22:414-427(1999).
CC -!- CATALYTIC ACTIVITY: 4-AMINO-BUTANOATE + 2-OXOGLUTARATE = SUCCINATE
CC SEMIALDEHYDE + L-GLUTAMATE.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -!- SUBUNIT: MONOMER (PROBABLE).
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -!- TISSUE SPECIFICITY: LIVER > PANCREAS > BRAIN > KIDNEY > HEART >
CC PLACENTA.
CC -!- DISEASE: DEFECTS IN ABAT ARE A CAUSE OF GABA-AT DEFICIENCY WHOSE
CC PHENOTYPE INCLUDES PSYCHOMOTOR RETARDATION, HYPOTONIA,
CC HYPERREFLEXIA, LETHARGY, REFRACTORY SEIZURES, AND EEG
CC ABNORMALITIES.
CC -!- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: L32961; AAA74449.1; -.
DR MIM: 137150; -.
DR InterPro: IPR000954; -.
DR Pfam: PF0202; aminotran_3; 1.
DR PROSITE: PS00600; AA_TRANSFER_CLASS.3; 1.
KW Transferase; Aminotransferase; Pyridoxal phosphate;
KW Neurotransmitter degradation; Mitochondrion; Transit peptide;
KW Disease mutation.
FT TRANSIT 1 28 MITOCHONDRION.
FT CHAIN 29 500 4-AMINO-BUTYRATE AMINOTRANSFERASE.
FT BINDING 357 357 PYRIDOXAL PHOSPHATE.
FT VARIANT 220 220 R -> K (IN GABA-AT DEFICIENCY; 25%
FT REDUCTION IN ACTIVITY).
FT /FTID=VAR_008883.
FT CONFLICT 109 109 D -> H (IN REF. 2).
FT CONFLICT 113 113 V -> L (IN REF. 2).
FT CONFLICT 132 132 E -> G (IN REF. 2).
FT CONFLICT 191 191 K -> Q (IN REF. 2).
FT CONFLICT 204 204 W -> G (IN REF. 2).
FT CONFLICT 216 216 S -> A (IN REF. 2).
FT CONFLICT 268 268 G -> R (IN REF. 2).
FT CONFLICT 320 320 C -> G (IN REF. 2).
FT CONFLICT 366 366 L -> H (IN REF. 2).
FT SEQUENCE 500 AA; 56557 MW; 41199085693F80AD CRC64;

Query Match 38.8%; Score 41.5; DB 1; Length 500;
Best Local Similarity 55.6%; Pred. No. 39;
Matches 10; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 3 TFOVEVPGSQHIDSQKA 20
Db 17 TYRLVPGSRHI-SQAAA 33

RESULT 13
RPH_ECOLI STANDARD; PRT; 141 AA.
ID RPH_ECOLI
AC P28369; P77246;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PEPTIDE CHAIN RELEASE FACTOR HOMOLOG (RF-H).
GN PRFH.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90330577; PubMed=1695895;
RA Henrich B., Monnerjahn U., Plapp R.;
RT "Peptidase D gene (pepD) of Escherichia coli K-12: nucleotide
RT sequence, transcript mapping, and comparison with other peptidase
RT genes.";
RL J. Bacteriol. 172:4641-4651(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R.,
RA Ichihara S., Mizuno T., Makino K., Nakata A., Yura T., Sempel G.,
RA Mizobuchi K.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [3]

```

RP SEQUENCE FROM N.A.  
RX STRAIN-K12 / MG1655;  
RA MEDLINE-97426617; PubMed-9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RL "The complete genome sequence of *Escherichia coli* K-12.";  
RL Science 277:1453-1474(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,  
RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,  
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,  
RA Davis R.W.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP IDENTIFICATION OF PROTEIN.  
RX MEDLINE-93027135; PubMed-1408743;  
RA Pel H.J., Rep M., Grivell L.A.;  
RT "Sequence comparison of new prokaryotic and mitochondrial members of  
RT the polypeptide chain release factor family predicts a five-domain  
RT model for release factor structure.";  
RL Nucleic Acids Res. 20:4423-4428(1992).  
CC -!- FUNCTION: COULD BE PROTEIN FACTOR INVOLVED IN TRANSLATIONAL  
CC TERMINATION.  
CC  
CC -!- SIMILARITY: BELONGS TO THE PROKARYOTIC AND MITOCHONDRIAL RELEASE  
CC FACTORS FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; M34034; -; NOT ANNOTATED\_CDS.  
DR EMBL; D83536; BAA77905.1; -;  
DR EMBL; AE000132; AAC73340.1; -;  
DR EMBL; U70214; AAB08656.1; -;  
DR Ecogene; EG11496; prfh.  
DR InterPro; IPR000352; -;  
DR Pfam; PF00472; RF-1; 1.  
DR PROSITE; PS00745; RF\_PROK\_I; 1.  
KW Protein biosynthesis.  
FT DOMAIN 104 107 POLY-GLN.  
FT CONFLICT 1 6 MGIRK -> MLETETGRYSDTLRSALVSLDGDNAWALSES  
FT (IN REF. 3 AND 4).  
FT  
SQ SEQUENCE 141 AA; 16177 MW; 9D2EBD9AF7A04831 CRC64;  
  
Query Match 38.3%; Score 41; DB 1; Length 141;  
Best Local Similarity 35.0%; Pred. No. 12;  
Matches 7; Conservative 4; Mismatches 9; Indels 0; Gaps 0;  
  
QY 2 ERFQVEVPGSQHDSQKKA1 21  
II : ||| :  
DB 49 ETLRSSGGGQHVKNKTSVA 68  
  
RESULT 14  
YP68\_CAEEL STANDARD; PRT; 313 AA.  
ID YP68\_CAEEL  
AC Q09217;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL 37.0 KDA PROTEIN B0495.8 IN CHROMOSOME II.  
GN B0495.8.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Kirsten J.;  
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: SOME, TO YEAST YDL087C AND S.POMBE SPOC16A11.13.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; U21317; AAG62527.1; -;  
DR WormPep; B0495.8; CE01766.  
KW Hypothetical protein.  
SQ SEQUENCE 313 AA; 36977 MW; 00D2327621BFED0 CRC64;  
  
Query Match 38.3%; Score 41; DB 1; Length 313;  
Best Local Similarity 46.2%; Pred. No. 28;  
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;  
  
QY 7 EYVPGSQHDSQK 19  
:: ||||| :  
DB 11 QLMGSHQVDNKEK 23  
  
RESULT 15  
TGT\_ZYMMO STANDARD; PRT; 385 AA.  
AC F28720; Q60247;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE QUEUINE TRNA-RIBOSYLTRANSFERASE (EC 2.4.2.29) (TRNA-GUANINE  
DE TRANSGLYCOSYLASE) (GUANINE INSERTION ENZYME).  
GN TGT.  
OS Zymomonas mobilis.  
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;  
OC Zymomonas.  
OX NCBI\_TaxID=542;  
RN [1]  
RP SEQUENCE FROM N.A., SEQUENCE OF 1-6, AND CHARACTERIZATION.  
RX MEDLINE-95394847; PubMed-7665516;  
RA Reuter K.K.H., Ficner R.;  
RT "Sequence analysis and overexpression of the Zymomonas mobilis tgt  
RT gene encoding TRNA-guanine transglycosylase: purification and  
RT biochemical characterization of the enzyme.";  
RL J. Bacteriol. 177:5284-5288(1995).  
RN [2]  
RP PARTIAL SEQUENCE FROM N.A.  
RC STRAIN-ATCC 31821 / ZM4 / CP4;  
RX MEDLINE-92406015; PubMed-15266462;  
RA Shark K.B., Conway T.;  
RT "Cloning and molecular characterization of the DNA ligase gene (lig)  
RT from Zymomonas mobilis.";  
RL FEMS Microbiol. Lett. 75:19-26(1992).  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS).  
RX MEDLINE-96256303; PubMed-8654383;  
RA Romier C., Reuter K., Suck D., Ficner R.;  
RT "Crystal structure of TRNA-guanine transglycosylase: RNA modification  
RT by base exchange.";  
RL EMBO J. 15:2850-2857(1996).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).  
RA Graedler U., Gerber H.D., Goodenough-Lashua D.M., Garcia G.A.G.,  
RA Ficner R., Reuter K., Stubbs M.T., Klebe G.;  
RT "A new target for shigellosis: rational design and crystallographic



```

RT studies of inhibitors of trna-guanine transglycosylase."
RL Submitted (MAR-2000) to the PDB data bank.
CC -|- FUNCTION: EXCHANGES THE GUANINE RESIDUE WITH 7-AMINOMETHYL-7-
CC DEAZAGUANINE IN TRNAS WITH GU(N) ANTICODONS (TRNA-ASP, -ASN, -HIS
CC AND -TYR). AFTER THIS EXCHANGE, A CYCLOPENTIDIOL MOIETY IS
CC ATTACHED TO THE 7-AMINOMETHYL GROUP OF 7-DEAZAGUANINE, RESULTING
CC IN THE HYPERMODIFIED NUCLEOSIDE QUEUOSINE (Q) (7-((4,5-CIS-
CC DIHYDROXY-2-CYCLOPENTEN-1-YL)AMINO)METHYL)-7-DEAZAGUANOSINE).
CC -|- CATALYTIC ACTIVITY: TRNA GUANINE + QUEUINE = TRNA QUEUINE +
CC GUANINE.
CC -|- COFACTOR: BINDS AND REQUIRES ZINC FOR ACTIVITY. ALSO REQUIRES
CC MAGNESIUM.
CC -|- SUBUNIT: MONOMER.
CC -|- SIMILARITY: BELONGS TO THE QUEUINE TRNA-RIBOSYLTRANSFERASE FAMILY.
CC -|- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN EXTENSIVELY DUE
CC TO FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L33777; AAA27704.1; ALT_INIT.
CC EMBL; L33777; AAA27705.1; ALT_INIT.
CC EMBL; Z11910; -; NOT_ANNOTATED_CDS.
CC DR PDB; 1PUD; 07-JUL-97.
CC DR PDB; 1WKD; 07-JUL-97.
CC DR PDB; 1WKE; 07-JUL-97.
CC DR PDB; 1WKF; 07-JUL-97.
CC DR PDB; 1WLF; 07-JUL-97.
CC DR PDB; 1WNG; 19-APR-00.
CC DR InterPro; IPR002616; -.
CC DR Pfam; PF01702; TGT. 1.
CC KW Queuosine biosynthesis; Transferase; Glycosyltransferase;
CC KW tRNA processing; zinc; Magnesium; 3D-structure.
CC FT INIT_MET 0 0
CC FT ACT_SITE 102 102
CC FT ACT_SITE 280 280
CC FT METAL 317 317 ZINC.
CC FT METAL 319 319 ZINC.
CC FT METAL 322 322 ZINC.
CC FT METAL 348 348 ZINC.
CC SQ SEQUENCE 385 AA; 42738 MW; F3D6FA270A0B23F3 CRC64;

Query Match 38.3%; Score 41; DB 1; Length 385;
Best Local Similarity 38.1%; Pred. No. 35;
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 GENFOVEVPGSQHIDSOKKAI 21
| | | : | | : | : | |
DB 120 GVTFKSHLDCSRHMLSPERSI 140

RESULT 16
YAAN_BACSU STANDARD; PRT; 386 AA.
AC P37535;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HYPOTHETICAL 43.8 KDA PROTEIN IN XPAC-ABRB INTERGENIC REGION.
GN YAAN.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
MEDLINE=96051385; PubMed=7584024;

Query Match 38.3%; Score 41; DB 1; Length 386;
Best Local Similarity 31.2%; Pred. No. 35;
Matches 5; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 6 VEYPGSQHIDSOKKAI 21
| | | | : | | : | : | |
DB 22 IEIPGSEAVRAKEQV 37

RESULT 17
RPC_BPPHC STANDARD; PRT; 683 AA.
AC P08979;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE REPRESSOR PROTEIN C.
GN C.
OS Bacteriophage phi-C31.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
OC Lambda phage group.
OX NCBI_TaxID=10719;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-NORWICH;
RX MEDLINE=89039715; PubMed=3185504;
RA Sinclair R.B., Bibb M.J.;
RT "The repressor gene (c) of the Streptomyces temperate phage phi c31:
RT nucleotide sequence, analysis and functional cloning.";
RL Mol. Gen. Genet. 213:269-277(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94374705; PubMed=8088546;
RA Hartley N.M., Murphy G.O., Bruton C.J., Chater K.F.;
RT "Sequence of the essential early region of phi C31, a temperate phage
RT of Streptomyces spp. with unusual features in its lytic
RT development.";
RL Gene 147:29-40(1994).
RN [2]
RP SEQUENCE FROM N.A.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RT subtilis chromosome containing the replication origin.";
RL DNA Res. 1:1-14(1994).
RN [2]
RP SEQUENCE OF 1-191 FROM N.A.
RA Bookstein C., Edwards C.W., Hulett F.M.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D26185; BAA05262.1; -.
CC EMBL; M96156; AAA22892.1; -.
CC EMBL; Z99104; CAB11802.1; -.
CC Subtilisin; BGL0090; yaan.
KW Hypothetical protein.
FT CONFLICT 191 191 E -> M (IN REF. 2).
SQ SEQUENCE 386 AA; 43830 MW; D921F3A0F6845EEE CRC64;

Query Match 38.3%; Score 41; DB 1; Length 386;
Best Local Similarity 31.2%; Pred. No. 35;
Matches 5; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 6 VEYPGSQHIDSOKKAI 21
| | | | : | | : | : | |
DB 22 IEIPGSEAVRAKEQV 37

RESULT 17
RPC_BPPHC STANDARD; PRT; 683 AA.
AC P08979;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE REPRESSOR PROTEIN C.
GN C.
OS Bacteriophage phi-C31.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
OC Lambda phage group.
OX NCBI_TaxID=10719;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-NORWICH;
RX MEDLINE=89039715; PubMed=3185504;
RA Sinclair R.B., Bibb M.J.;
RT "The repressor gene (c) of the Streptomyces temperate phage phi c31:
RT nucleotide sequence, analysis and functional cloning.";
RL Mol. Gen. Genet. 213:269-277(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94374705; PubMed=8088546;
RA Hartley N.M., Murphy G.O., Bruton C.J., Chater K.F.;
RT "Sequence of the essential early region of phi C31, a temperate phage
RT of Streptomyces spp. with unusual features in its lytic
RT development.";
RL Gene 147:29-40(1994).
RN [2]
RP SEQUENCE FROM N.A.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

DR EMBL; X12865; CAA31345.1; -;  
 DR EMBL; X76288; CAA53911.1; -;  
 DR PIR; S01433; S01433.  
 KW Transcription regulation; Repressor; DNA-binding.  
 SQ SEQUENCE 683 AA; 74077 MW; B02379D204F37D1B CRC64;

Query Match 38.3%; Score 41; DB 1; Length 683;  
 Best Local Similarity 50.0%; Pred. No. 65;  
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSOHI 14  
 | | | | | | | | | |  
 DB 86 GRTFDTLTGADHI 99

RESULT 18  
 DPOL\_THELI STANDARD; PRT; 1702 AA.  
 AC P30317;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE DNA POLYMERASE (EC 2.7.7.7) (VENT DNA POLYMERASE) [CONTAINS: PI-TLI I  
 DE ENDONUCLEASE; PI-TLI II ENDONUCLEASE].  
 GN POL.  
 OS Thermococcus litoralis.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.  
 OX NCBI\_TaxID=2265;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=92302285; PubMed=1608969;  
 RA Perler F.B., Comb D.G., Jack W.E., Moran L.S., Qiang B.,  
 RA Kucera R.B., Benner J., Slatko B.E., Nwankwo D.O., Hempstead S.K.,  
 RA Carlow C.K.S., Jannasch H.;  
 RT "Intervening sequences in an Archaea DNA polymerase gene."  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:5577-5581(1992).  
 RN [2]  
 RP PROTEIN SPLICING.  
 RX MEDLINE=93117083; PubMed=1475179;  
 RA Hodges R.A., Perler F.B., Noren C.J., Jack W.E.;  
 RT "Protein splicing removes intervening sequences in an archaea DNA  
 RT polymerase."  
 RL Nucleic Acids Res. 20:6153-6157(1992).  
 CC -!- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE  
 CC EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.  
 CC -!- FUNCTION: INTEIN ENCODED ENDONUCLEASES ARE THOUGHT TO MEDIATE  
 CC INTEIN MOBILITY BY SITE-SPECIFIC RECOMBINATION INITIATED BY  
 CC ENDONUCLEASE CLEAVAGE AT THE "HOMING SITE" IN GENE THAT LACK THE  
 CC INTEIN.  
 CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =  
 CC N PYROPHOSPHATE + DNA(N).  
 CC -!- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES  
 CC A POST-TRANSLATIONAL EXCISION OF THE TWO INTERVENING REGION  
 CC (INTEINS) FOLLOWED BY PEPTIDE LIGATION.  
 CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M74198; AAA72100.1; -;  
 DR EMBL; M74198; AAA72101.1; -;  
 DR PIR; S42459; S42459.  
 DR REBASE; 2613; PI-Tlii.  
 DR REBASE; 2621; PI-Tlii.  
 DR InterPro; IPR002064; -;  
 DR InterPro; IPR002203; -;  
 DR Pfam; PF00136; DNA\_pol\_B; 3.

DR PRINTS; PR00379; INTEIN.  
 DR PROSITE; PS00116; DNA\_POLYMERASE\_B; FALSE\_NEG.  
 DR PROSITE; PS00881; PROTEIN\_SPLICING; 2.  
 KW Transferase; DNA-directed DNA polymerase; DNA replication;  
 KW DNA-binding; Hydrolase; Endonuclease; Autocatalytic cleavage;  
 KW Protein splicing.  
 FT CHAIN 1 494 DNA POLYMERASE, 1ST PART.  
 FT CHAIN 495 1032 PI-TLI II ENDONUCLEASE (TLI POL-1 INTEIN)  
 FT (IVPS1).  
 FT CHAIN 1033 1081 DNA POLYMERASE, 2ND PART.  
 FT CHAIN 1082 1471 PI-TLI I ENDONUCLEASE (TLI POL-2 INTEIN)  
 FT (IVPS2).  
 FT CHAIN 1472 1702 DNA POLYMERASE, 3RD PART.  
 FT CHAIN 1703 197293 MW; 21D6B98C75F53B20 CRC64;  
 SQ SEQUENCE 1702 AA; 197293 MW; 21D6B98C75F53B20 CRC64;

Query Match 38.3%; Score 41; DB 1; Length 1702;  
 Best Local Similarity 46.7%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSOHID 15  
 | | | | | | | | | |  
 DB 247 GDSFAVEIKGRHFD 261

RESULT 19  
 DPOL\_THEST STANDARD; PRT; 1829 AA.  
 AC Q33845;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE DNA POLYMERASE (EC 2.7.7.7).  
 GN POL.  
 OS Thermococcus sp. (strain TY).  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.  
 OX NCBI\_TaxID=86030;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98094267; PubMed=9434178;  
 RA Niehaus F., Frey B., Antranikian G.;  
 RT "Cloning and characterisation of a thermostable alpha-DNA polymerase  
 RT from the hyperthermophilic archaeon Thermococcus sp. TY."  
 RL Gene 204:153-158(1997).  
 CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =  
 CC N PYROPHOSPHATE + DNA(N).  
 CC -!- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES  
 CC A POST-TRANSLATIONAL EXCISION OF THE THREE INTERVENING REGION  
 CC (INTEINS) FOLLOWED BY PEPTIDE LIGATION.  
 CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Y13030; CAA73475.1; -;  
 DR InterPro; IPR002064; -;  
 DR InterPro; IPR002203; -;  
 DR Pfam; PF00136; DNA\_pol\_B; 4.  
 DR PRINTS; PR00379; INTEIN  
 DR PROSITE; PS00116; DNA\_POLYMERASE\_B; FALSE\_NEG.  
 DR PROSITE; PS00881; PROTEIN\_SPLICING; 3.  
 KW Transferase; DNA-directed DNA polymerase; DNA replication;  
 KW DNA-binding; Hydrolase; Endonuclease; Autocatalytic cleavage;  
 KW Protein splicing.  
 FT CHAIN 1 409 DNA POLYMERASE, 1ST PART (POTENTIAL).  
 FT CHAIN 410 769 INTEIN I.  
 FT CHAIN 770 855 DNA POLYMERASE, 2ND PART (POTENTIAL).  
 FT CHAIN 856 1392 INTEIN II.

```
FT CHAIN 1393 1441 DNA POLYMERASE, 3RD PART (POTENTIAL).
FT CHAIN 1442 1598 INTEIN III.
FT CHAIN 1599 1829 DNA POLYMERASE, 4TH PART (POTENTIAL).
SQ SEQUENCE 1829 AA; 211875 MW; A113A8BC57E9CB3 CRC64;

Query Match 38.3%; Score 41; DB 1; Length 1829;
Best Local Similarity 46.7%; Pred. No. 1.9e+02;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSQHID 15
   |::| | | | | |
DB 247 GDSFAVEIKGRIHFD 261

RESULT 20
PMS2_HUMAN STANDARD; PRT; 862 AA.
ID PMS2_HUMAN
AC P54278;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PMS1 PROTEIN HOMOLOG 2 (DNA MISMATCH REPAIR PROTEIN PMS2).
GN PMS2 OR PMSL2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Endometrial tumor;
RX MEDLINE=94352394; PubMed=8072530;
RA Nicolaides N.C., Papadopoulos N., Liu B., Wei Y.-F., Carter K.C.,
RA Ruben S.M., Rosen C.A., Haseltine W.H., Fleischmann R.D.,
RA Fraser C.M., Adams M.D., Venter J.C., Dunlop M.G., Hamilton S.R.,
RA Petersen G.M., de la Chapelle A., Vogelstein B., Kinzler K.W.;
RT "Mutations of two PMS homologues in hereditary nonpolyposis colon
RT cancer.";
RL Nature 371:75-80(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Bronner C.E., Baker S.M., Morrison P.T., Warren G., Smith L.G.,
RA Lescoe M.K., Kane M., Earlbino C., Lipford J., Lindblom A.,
RA Tannergaard P., Bollag R.J., Godwin A.R., Ward D.C.,
RA Nordenskjold M., Fishel R., Kolodner R.D., Liskay R.M.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLY INVOLVED IN THE REPAIR OF MISMATCHES IN DNA.
CC -1- SUBUNIT: HETERODIMER OF PMS2 AND MLH1.
CC -1- DISEASE: ASSOCIATED WITH FAMILIAL HEREDITARY NONPOLYPOSIS COLON
CC CANCER (HNPCC) (LYNCH SYNDROME). HNPCC IS ONE OF THE MOST COMMON
CC GENETIC DISEASES IN THE WESTERN WORLD, AND ACCOUNTS FOR 15% OF ALL
CC COLON CANCERS.
CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTL/HEXB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U13696; AAA63923.1; -
CC EMBL; U14658; AAA50390.1; -
CC DR SWISS-2DPAGE; P54278; HUMAN.
CC MIM; 600259; -
CC DR InterPro; IPR002099; -
CC DR Pfam; PF01119; DNA_mis_repair; 1.
CC DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
CC KW DNA repair; Disease mutation; Anti-oncogene; Polymorphism.
FT VARIANT 20 20 R -> Q.
FT FTID=VAR_004469,
FT FTID=VAR_004469,
FT CONFLICT 470 470 P -> S (IN REF. 2).
```

```
SQ SEQUENCE 862 AA; 95797 MW; B60A605222CBBAC CRC64;

Query Match 37.9%; Score 40.5; DB 1; Length 862;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 9; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

OY 10 GSQ-HIDSQKKA 20
   | | | | | | | | | |
DB 528 GSQEHVDSQKKA 539

RESULT 21
UBIQ_NEUCR STANDARD; PRT; 76 AA.
ID UBIQ_NEUCR
AC P13117;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE UBIQUITIN.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC NCBI_TaxID=5141;
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ST. LAWRENCE 74 / SL 74 / ORS 6A;
RX MEDLINE=89366647; PubMed=2549509;
RA Taccioli G.E., Grotewold E., Aisemberg G.O., Judewicz N.D.;
RT "Ubiquitin expression in Neurospora crassa: cloning and sequencing of
RT a polyubiquitin gene.";
RL Nucleic Acids Res. 17:6153-6165(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ST. LAWRENCE 74 / SL 74 / ORS 6A;
RX MEDLINE=91327370; PubMed=1650731;
RA Taccioli G.E., Grotewold E., Aisemberg G.O., Judewicz D.N.;
RT "The cDNA sequence and expression of an ubiquitin-tail gene fusion in
RT Neurospora crassa.";
RL Gene 102:133-137(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RX MEDLINE=94374698; PubMed=8088539;
RA Tarawneh K.A., Anumula K.R., Free S.J.;
RT "The isolation and characterization of a Neurospora crassa gene
RT (ubi::crp-6) encoding a ubiquitin-40S ribosomal protein fusion
RT protein.";
RL Gene 147:137-140(1994).
CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF
CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE
CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME
CC BIOGENESIS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.
CC -1- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN
CC PRECURSOR WITH 4 EXACT HEAD TO TAIL REPEATS. THERE IS A FINAL
CC AMINO-ACID (GLN) AFTER THE LAST REPEAT. SOME UBIQUITIN GENES
CC CONTAIN A SINGLE COPY OF UBIQUITIN FUSED TO A RIBOSOMAL PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X13140; CAA31530.1; ALT_TERM.
CC EMBL; U01220; AAA56880.1; ALT_TERM.
CC DR EMBL; U01220; AAA03351.1; ALT_TERM.
CC DR PIR; S05323; UONC.
CC DR HSP; P02248; 1AAR.
CC InterPro; IPR000626; -.
```

```
DR Pfam: PF00240; ubiquitin; 1.
DR PRINTS: PRO0348; UBIQUITIN.
DR PROSITE: PS00299; UBIQUITIN_1; 1.
DR PROSITE: PS0053; UBIQUITIN_2; 1.
KW Nuclear protein; Polyprotein.
FT SITE 48 48
FT BINDING 76 76
SQ SEQUENCE 76 AA; 8598 MW; 994480FE7D38403E CRC64;

NECESSARY FOR BRANCHED-CHAIN
MULTIUBIQUITIN ADDUCTS.
CONJUGATION TO ACCEPTOR PROTEINS.
994480FE7D38403E CRC64;

Query Match 37.4%; Score 40; DB 1; Length 76;
Best Local Similarity 42.9%; Pred. No. 9;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHIDSOKKAI 21
   I:::||| ||| ||| |||
DB 10 GKTILEVESDIDNVKQKI 30

RESULT 22
ID IF5A_AERPE STANDARD; PRT; 148 AA.
AC Q9YA53.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSLATION INITIATION FACTOR 5A (EIF-5A) (HYPUSINE-CONTAINING
  PROTEIN).
GN EIF5A OR APE2085.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
  crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 5:83-101(1999).
CC -!- FUNCTION: FUNCTIONS BY PROMOTING THE FORMATION OF THE FIRST
CC PEPTIDE BOND (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE EIF-5A FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AP000063; BAA81096.1; -.
DR InterPro: IPR001884; -.
DR Pfam: PF01287; eIF-5a; 1.
DR PROSITE: PS00302; IF5A_HYPUSINE; 1.
KW Protein biosynthesis; Initiation factor; Hypusine.
FT MOD_RES 50 50
SQ SEQUENCE 148 AA; 16205 MW; B7C986047F0FB46F CRC64;

Query Match 37.4%; Score 40; DB 1; Length 148;
Best Local Similarity 42.1%; Pred. No. 18;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
```

```
QY 2 ETQVEVPGSQHIDSOKKA 20
   ::||| ||| ||| |||
DB 103 DTEVEKPGGNEEEQLAA 121

RESULT 23
HS74_CAEEL STANDARD; PRT; 288 AA.
ID HS74_CAEEL
AC P20163.
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HEAT SHOCK 70 KDA PROTEIN D (FRAGMENT).
GN HSP-4 OR HSP70D.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89306577; PubMed=2744444;
RA Heschl M.F.P., Baillie D.L.;
RT "Identification of a heat-shock pseudogene from Caenorhabditis
  elegans.";
RL Genome 32:190-195(1989).
CC -!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY. STRONG,
CC TO MAMMALIAN GRP78.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: M28528; AAA28076.1; -.
DR HSP: P19120; IATS.
DR InterPro: IPR000886; -.
DR InterPro: IPR001023; -.
DR Pfam: PF00012; HSP70; 1.
DR PROSITE: PS00014; ER_TARGET; 1.
DR PROSITE: PS00297; HSP70_1; PARTIAL.
DR PROSITE: PS00329; HSP70_2; PARTIAL.
DR PROSITE: PS01036; HSP70_3; PARTIAL.
KW ATP-binding; Heat shock; Endoplasmic reticulum; Multigene family.
FT SITE 285 288
FT NON_TER 1
FT SEQUENCE 288 AA; 31267 MW; 967F5A4A12FA67BF CRC64;
SQ SEQUENCE 288 AA; 31267 MW; 967F5A4A12FA67BF CRC64;

Query Match 37.4%; Score 40; DB 1; Length 288;
Best Local Similarity 35.3%; Pred. No. 37;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 3 TFQVEVPGSQHIDSOKK 19
   ||:::||| ||::||
DB 133 TFEIDVNGILHVSADK 149

RESULT 24
YRYL_CAEEL STANDARD; PRT; 355 AA.
ID YRYL_CAEEL
AC Q10005;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 39.9 KDA PROTEIN T15H9.1 IN CHROMOSOME II PRECURSOR.
GN T15H9.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
```

```

OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Gardner A.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 J DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z47356; CAA87414.1; -
CC DR HSSP; P08622; LXBL.
CC DR WormPep; T15H9.1; CE01664.
CC DR InterPro; IPR001623; -
CC DR InterPro; IPR002939; -
CC DR Pfam; PF00226; DnaJ; 1.
CC DR Pfam; PF01556; DnaJ_C; 1.
CC DR PROSITE; PS00636; DnaJ_1; 1.
CC DR PROSITE; PS00076; DnaJ_2; 1.
CC DR Hypothetical protein; Chaperone; Signal.
KW SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 355 HYPOTHETICAL PROTEIN T15H9.1.
FT DOMAIN 22 91 J-DOMAIN.
FT DOMAIN 97 103 POLY-GLY.
FT DOMAIN 116 123 POLY-GLY.
FT SEQUENCE 355 AA; 39861 MW; 95FA4D8E551D9CC2 CRC64;
SQ
Query Match 37.4%; Score 40; DB 1; Length 355;
Best Local Similarity 47.4%; Pred. No. 47;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;
QY 3 TFOVEVPGSOHIDSQKAI 21
DB 320 TFDVEFPKTELSDEQRAQI 338
RESULT 25
ADH3_ENTHI STANDARD; PRT; 395 AA.
AC Q24857; Q94560;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALCOHOL DEHYDROGENASE 3 (EC 1.1.1.1) (ADH).
GN ADH3.
OS Entamoeba histolytica.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=5759;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HM-1:IMSS;
RX MEDLINE=96201701; PubMed=8705667;
RA Kimura A., Hara Y., Kimoto T., Okuno Y., Minekawa Y.,
RA Nakabayashi T.;
RT "Cloning and expression of a putative alcohol dehydrogenase gene of
RT Entamoeba histolytica and its application to immunological
RT examination."
RL Clin. Diagn. Lab. Immunol. 3:270-274(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HM-1:IMSS;
RX MEDLINE=96201701; PubMed=8611619;
RA Rodriguez M.A., Baez-Camargo M., Delgadillo D.M., Orozco E.;
RT "Cloning and expression of an Entamoeba histolytica NAPD+(-)dependent
RT alcohol dehydrogenase gene."
RL Biochim. Biophys. Acta 1306:23-26(1996).
```

```

CC -!- FUNCTION: NAPD+(-)DEPENDENT ADH ACTIVITY.
CC -!- CATALYTIC ACTIVITY: AN ALCOHOL + NADP(+) = AN ALDEHYDE OR KETONE +
CC NADPH.
CC -!- SIMILARITY: BELONGS TO C.ACETOBUTYLICUM BDHA AND BDHB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D49910; BAA08651.1; -
CC DR EMBL; Z48752; CAA88639.1; -
CC DR InterPro; IPR001670; -
CC DR Pfam; PF00465; Fe-ADH; 2.
CC DR PROSITE; PS00913; ADH_IRON_1; FALSE_NEG.
CC DR PROSITE; PS00060; ADH_IRON_2; 1.
CC KW Oxidoreductase; NADP.
CC FT CONFLICT 193 193 R -> I (IN REF. 2).
CC FT CONFLICT 235 242 TLENHODY -> HKKIIRTI (IN REF. 2).
CC FT CONFLICT 273 283 MISSING (IN REF. 2).
CC FT CONFLICT 307 309 VMR -> NE (IN REF. 2).
CC FT SEQUENCE 395 AA; 43485 MW; 506FBA933631FA9 CRC64;
SQ
Query Match 37.4%; Score 40; DB 1; Length 395;
Best Local Similarity 35.0%; Pred. No. 52;
Matches 7; Conservative 3; Mismatches 10; Indels 0; Gaps 0;
QY 1 GETFQVEVPGSOHIDSQKKA 20
DB 297 GETLAITPGVMRFNKEKNA 316
Search completed: July 16, 2001, 16:44:05
Job time: 464 sec
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 16:43:38 ; Search time 57.41 Seconds  
(without alignments)  
48.396 Million cell updates/sec

Title: US-09-786-648-5  
Perfect score: 107  
Sequence: 1 GETFQVEVPSQSHDSQKKAI 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_16.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_unclassified.\*
- 13: sp\_vertebrate.\*
- 14: sp\_virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	94.4	124	2 Q57193	Q57193 vibrio chol
2	101	94.4	124	2 Q56635	Q56635 vibrio chol
3	101	94.4	124	2 Q9RE15	Q9RE15 vibrio chol
4	95	88.8	103	2 Q9R646	Q9R646 vibrio chol
5	54	50.5	508	4 Q99987	Q99987 homo sapien
6	52	48.6	552	10 Q9LJB0	Q9LJB0 arabidopsis
7	49	45.8	428	10 Q9LGM2	Q9LGM2 oryza sativ
8	47	43.9	392	10 Q40742	Q40742 oryza sativ
9	47	43.9	395	5 Q9NKD5	Q9NKD5 drosophila
10	46	43.0	91	10 Q9SYF4	Q9SYF4 arabidopsis
11	46	43.0	574	10 Q9SXC9	Q9SXC9 arabidopsis
12	45	42.5	1166	3 Q9P4E0	Q9P4E0 ustilago ma
13	45	42.1	439	10 Q9LY67	Q9LY67 arabidopsis
14	45	42.1	2732	14 Q9J3F2	Q9J3F2 murine hepa
15	45	42.1	2733	14 Q39226	Q39226 murine hepa
16	45	42.1	2733	14 Q9PYA2	Q9PYA2 murine hepa
17	45	42.1	2733	14 Q9J3E8	Q9J3E8 murine hepa
18	44	41.6	565	10 Q22511	Q22511 vitis vinif
19	44	41.1	293	10 Q49876	Q49876 lupinus alb

20	44	41.1	330	11 Q9OZL7	Q9OZL7 mus musculu
21	44	41.1	374	5 Q9U3M9	Q9U3M9 caenorhabdi
22	44	41.1	467	6 Q9N136	Q9N136 ovis aries
23	44	41.1	467	11 P97431	P97431 mus musculu
24	44	41.1	864	3 Q9P525	Q9P525 neurospora
25	43	40.7	641	10 Q9SKB2	Q9SKB2 arabidopsis
26	43	40.2	356	5 Q9NF90	Q9NF90 leishmania
27	43	40.2	368	10 Q9SN44	Q9SN44 arabidopsis
28	43	40.2	843	14 Q9OMN7	Q9OMN7 hepatitis b
29	43	40.2	1714	13 Q90995	Q90995 gallus gall
30	43	40.2	1810	13 Q90824	Q90824 gallus gall
31	42	39.3	204	2 Q9HTA0	Q9HTA0 pseudomonas
32	42	39.3	325	4 Q9HB36	Q9HB36 homo sapien
33	42	39.3	427	13 Q42099	Q42099 cyprinus ca
34	42	39.3	855	4 Q9HCA3	Q9HCA3 homo sapien
35	42	39.3	855	4 Q9H3S0	Q9H3S0 homo sapien
36	42	39.3	1742	2 Q55583	Q55583 synechocyst
37	41	38.8	254	2 Q31335	Q31335 bacillus ce
38	41	38.8	311	11 Q9EPB5	Q9EPB5 mus musculu
39	41	38.3	246	5 Q9V9U6	Q9V9U6 drosophila
40	41	38.3	260	10 Q9MA16	Q9MA16 arabidopsis
41	41	38.3	269	2 Q70057	Q70057 bordetella
42	41	38.3	298	10 Q43098	Q43098 psophocarpu
43	41	38.3	399	2 Q9F5L7	Q9F5L7 zymomonas m
44	41	38.3	550	10 Q04636	Q04636 arabidopsis
45	41	38.3	551	10 Q9SQ79	Q9SQ79 pinus taeda

ALIGNMENTS

RESULT 1  
Q57193 ID Q57193 PRELIMINARY; PRT; 124 AA.  
AC Q57193;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE CHOLERA TOXIN B PROTEIN (CTB) PRECURSOR (CTB).  
GN CTXB.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=566;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLASSICAL STRAIN 569B;  
RX MEDLINE=91355224; PubMed=1883840;  
RA Dams E., De Wolf M., Dierick W.;  
RT "Nucleotide sequence analysis of the CT operon of the Vibrio cholerae  
RT classical strain 569B.";  
RL Biochim. Biophys. Acta 1090:139-141(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLASSICAL BIOTYPE 569B;  
RA Shi C., Cao C., Zhang J., Ma Q.;  
RL Chin. Biochem. J. 9:395-399(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLASSICAL BIOTYPE 569B;  
RA Xu L.;  
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X58785; CAA41591.1; -;  
DR EMBL; U25679; AAC34728.1; -;  
DR EMBL; A00931; CAA00098.1; -;  
DR HSSP; P01556; 2CHB.  
DR InterPro; IPR001835; -;  
DR Pfam; PF01376; Enterotoxin\_B; 1.  
DR PRINTS; PR00772; ENTEROTOXIN.  
DR ProDom; PD012805; -; 1.  
KW Signal.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 124 CHOLERA TOXIN B PROTEIN (CTB).  
SQ SEQUENCE 124 AA; 13919 MW; D6BF83FFF7924EA3 CRC64;

2  
:  
:



```

DR InterPro: IPR002290; -.
DR Pfam: PF00069; pkinase: 2.
DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR SMART: SM00220; S_TKc; 1.
KW ATP-binding; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 508 AA; 58126 MW; 157FBF8F48511AF4 CRC64;

Query Match 50.5%; Score 54; DB 4; Length 508;
Best Local Similarity 50.0%; Pred. No. 1.4;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 GETFQVEVPGSQHIDSQKAI 20
   |::|::|::|::|::|::|
Db 329 GQINVHTPNKQVDSQKAA 348

RESULT 6
Q9LJB0 PRELIMINARY; PRT; 552 AA.
AC Q9LJB0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE EMB|CAB77996.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones."
RL DNA Res. 7:217-221(2000).
DR EMBL: AP000606; BAB01195.1; -.
SQ SEQUENCE 552 AA; 63036 MW; 740EA16CDEBB2447 CRC64;

Query Match 48.6%; Score 52; DB 10; Length 552;
Best Local Similarity 58.8%; Pred. No. 3.2;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 QVEVPGSQHIDSQKAI 21
   |||::|::|::|::|
Db 147 QVEIPASMEIDETKAI 163

RESULT 7
Q9LGM2 PRELIMINARY; PRT; 428 AA.
AC Q9LGM2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-NAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ESTS AU056822(S20908).
OS Oryza sativa (Rice)
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
OC Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0041E11.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0433F09.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP002521; BAA96762.1; -.
DR EMBL: AP002539; BAB08201.1; -.
DR InterPro: IPR001552; -.
DR Pfam: PF00441; Acyl-CoA_dh; 1.
DR PROSITE: PS00073; ACYL_COA_DH_2; UNKNOWN_1.
SQ SEQUENCE 428 AA; 46132 MW; 8D34E3698A8E6367 CRC64;

Query Match 45.8%; Score 49; DB 10; Length 428;
Best Local Similarity 52.6%; Pred. No. 7.4;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 TFOVEVPGSQHIDSQKAI 21
   |::|::|::|::|::|
Db 183 TTATKVPGGWHIDGQRWI 201

RESULT 8
Q40742 PRELIMINARY; PRT; 392 AA.
ID Q40742;
AC Q40742;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NAR-2001 (TrEMBLrel. 16, Last annotation update)
DE OSRAD23.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
OC Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NIPPONBARE;
RX MEDLINE=97369378; PubMed=9225866;
RA Schultz T.F., Quatrano R.S.;
RT "Characterization and expression of a rice RAD23 gene.";
RL Plant Mol. Biol. 34:557-562(1997).
DR EMBL: U63530; AAB65841.1; -.
DR HSSP: P02248; ITBE.
DR Mendel; 15760; Oryza; 2927; 15760.
DR InterPro: IPR000449; -.
DR InterPro: IPR000626; -.
DR Pfam: PF00240; ubiquitin; 1.
DR Pfam: PF00627; UBA; 2.
DR PROSITE: PS00053; UBIQUITIN_2; 1.
DR SMART: SM00165; UBA; 1.
SQ SEQUENCE 392 AA; 41753 MW; BD6E08574CC7CACB CRC64;

Query Match 43.9%; Score 47; DB 10; Length 392;
Best Local Similarity 42.9%; Pred. No. 14;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 GETFQVEVPGSQHIDSQKAI 21
   |::|::|::|::|::|
Db 10 GSTFQIEVDSAQKADVAKRII 30

RESULT 9
Q9NKD5 PRELIMINARY; PRT; 395 AA.
ID Q9NKD5

```

AC Q9NKK5;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE HYPOTHETICAL 43.6 KDA PROTEIN.  
GN BG:DS01514.3.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidae; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Y, CN BW SP;  
RX MEDLINE=99403001; PubMed=10471707;  
RA Ashburner M., Misra S., Rote J., Lewis S.E., Blazej R., Davis T.,  
RA Doyle C., Galle R., George K., Harris N., Hartzell G., Harvey D.,  
RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,  
RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,  
RA Celnik S., Rubin G.M.;  
RT "An exploration of the sequence of a 2.9-Mb region of the genome of  
RT Drosophila melanogaster: the Adh region.";  
RL Genetics 153:179-219(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Y, CN BW SP;  
RA Celnik S.E., Agbavani A., Arcaina T.T., Baxter E., Blazej R.G.,  
RA Burenhoff C., Champ M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,  
RA Farfan D.E., Galle R., George R.A., Harris N.B., Hoskins R.A.,  
RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,  
RA Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,  
RA Nixon K., Pacle J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,  
RA Sethi H., Snir E., Svirska R.R., Wan K.H., Weinburg T., Zhang R.,  
RA Zieran L.L., Rubin G.M.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003408; AAF44847.1; -;  
DR Flybase; FBgn0028907; BG:DS01514.3.  
KW Hypothetical protein.  
SQ SEQUENCE 395 AA; 43561 MW; AE4F1CC4ADD3DA73 CRC64;

Query Match 43.9%; Score 47; DB 5; Length 395;  
Best Local Similarity 56.2%; Pred. No. 14;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 QVEVPGSQHDSOKKA 20  
Db 329 RVSPGSGTHADANA 344  
:|||||:|

RESULT 10  
Q9SYF4 PRELIMINARY; PRT; 91 AA.  
ID Q9SYF4  
AC Q9SYF4;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)  
DE F1511.6 PROTEIN.  
GN F1511.6.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV, COLUMBIA;  
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,  
RA Li J., Kremenetskaia I., Luros J., Ngan i., Gonzalez A., Altafi H.,  
RA Araujo R., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,  
RA Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,  
RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;  
RT "Arabidopsis thaliana chromosome 1 BAC F1511 sequence.";

RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006577; AAD25769.1; -;  
DR InterPro; IPR000626; -;  
DR Pfam; PF00240; ubiquitin; 2.  
SQ SEQUENCE 91 AA; 10142 MW; E8766823D6450267 CRC64;

Query Match 43.0%; Score 46; DB 10; Length 91;  
Best Local Similarity 47.6%; Pred. No. 4.3;  
Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 GETFQVEVPGSQHDSOKKAI 21  
Db 58 GKTFNLEVGSEIIQQKNMI 78  
:|||||:|

RESULT 11  
Q9SXC9 PRELIMINARY; PRT; 574 AA.  
ID Q9SXC9  
AC Q9SXC9;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)  
DE T17H3.2 PROTEIN.  
GN T17H3.2.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV, COLUMBIA;  
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,  
RA Lee J.M., Liu A., Li J., Gonzalez A., Liu K., Vaysberg M., Sakano H.,  
RA Chin C., Hwang B., Choi E., Chlou J., Altafi H., Araujo R., Brooks S.,  
RA Buehler E., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,  
RA Huizar L., Khan S., Kim C., Palm C., Rowley D., Shinn P., Walker M.,  
RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;  
RT "Arabidopsis thaliana chromosome 1 BAC T17H3 sequence.";

RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC005916; AAD45990.1; -;  
DR InterPro; IPR001382; -;  
DR Pfam; PF01532; Glyco\_hydro.47; 1.  
DR PRINTS; PR00747; GLYHDLASE47.  
SQ SEQUENCE 574 AA; 65707 MW; 4ACC456DE487EA93 CRC64;

Query Match 43.0%; Score 46; DB 10; Length 574;  
Best Local Similarity 58.8%; Pred. No. 31;  
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 TTFQVEVPGSQHDSOKK 19  
Db 19 TTFVWDPPSQHIEVRKK 35  
:|:|:|:|:|:|

RESULT 12  
Q9P4E0 PRELIMINARY; PRT; 1166 AA.  
ID Q9P4E0  
AC Q9P4E0;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE PHEROMONE-RESPONSIVE MAPKK KINASE UBC4.  
GN UBC4.  
OS Ustilago maydis (Smut fungus).  
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;  
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.  
OX NCBI\_TaxID=5270;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1/2;

```
QY 1 GETFQVEVPGSQHIDSQKKA 21
    | : : | | | | | :
Db 91 GKTYSMEGPGIQCDEHNKGL 111
```

RESULT	16
Q9PYA2	

ID Q9PYA2 PRELIMINARY; PRT; 2733 AA.  
AC Q9PYA2;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE RNA-DIRECTED RNA POLYMERASE.  
GN ORF1B.  
OS murine hepatitis virus, and murine hepatitis virus strain ML-11.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
OC Coronaviridae; Coronavirus.  
OX NCBI\_TaxID=11138, 123595;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=mouse hepatitis virus strain 2; STRAIN=MHV-2;  
RA Das Sarma J., Hingley S.T., Lai M.M.C., Weiss S.R., Lavi E.;  
RT "Pathogenesis and sequence analysis of mouse hepatitis virus type 2:  
RT an experimental model system of acute meningitis and hepatitis in  
RT mice.";  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=mouse hepatitis virus strain ML-11; STRAIN=ML-11;  
RA Das Sarma J., Hingley S.T., Lai M.M.C., Weiss S.R., Lavi E.;  
RT "Pathogenesis and sequence analysis of mouse hepatitis virus type 2:  
RT an experimental model system of acute meningitis and hepatitis in  
RT mice.";  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF201929; AAF19384.1; -  
DR EMBL; AF207902; AAF68920.1; -  
KW RNA-directed RNA polymerase.  
SQ SEQUENCE 2733 AA; 309420 MW; D04F6457578EC17 CRC64;

Query Match 42.1%; Score 45; DB 14; Length 2733;  
Best Local Similarity 52.9%; Pred. No. 2.4e+02;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 ETFQVPEVPSQHDSQK 18  
||||| || ||| :  
Db 1198 ETFQNNVNYQHIGMKR 1214

RESULT 17  
Q9J3E8 PRELIMINARY; PRT; 2733 AA.  
ID Q9J3E8  
AC Q9J3E8;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE RNA-DIRECTED RNA POLYMERASE.  
OS murine hepatitis virus  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
OC Coronaviridae; Coronavirus.  
OX NCBI\_TaxID=11138;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ML-10;  
RA Das Sarma J., Hingley S.T., Lai M.M.C., Weiss S.R., Lavi E.;  
RT "Pathogenesis and sequence analysis of mouse hepatitis virus type 2:  
RT an experimental model system of acute meningitis and hepatitis in  
RT mice.";  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF208067; AAF69342.1; -  
KW RNA-directed RNA polymerase.  
SQ SEQUENCE 2733 AA; 309251 MW; 47049486732FBAFE CRC64;

Query Match 42.1%; Score 45; DB 14; Length 2733;  
Best Local Similarity 52.9%; Pred. No. 2.4e+02;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 ETFQVPEVPSQHDSQK 18  
||||| || ||| :;

Db 1198 ETFQNNVNYQHIGMKR 1214  
RESULT 18  
O22511 PRELIMINARY; PRT; 565 AA.  
ID O22511  
AC O22511;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE GLUTATHIONE REDUCTASE (NADPH) (EC 1.6.4.2) (FRAGMENT).  
GN GOR.  
OS Vitis vinifera (Grape).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Vitaceae; Vitis.  
OX NCBI\_TaxID=29760;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. THOMPSON SEEDLESS (CLONE 2A); TISSUE=FRUIT;  
RA Cassol T., Adams D.O.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- COFACTOR: FAD (BY SIMILARITY).  
CC -!- SIMILARITY: TO PYRIDINE NUCLEOTIDE-DISULPHIDE OXIDOREDUCTASES  
CC CLASS-I.  
DR EMBL; AF019907; AAB70837.1; -  
DR HSSP; P00390; IALG.  
DR Mendel; 26381; Vitvi; 1190; 26381.  
DR InterPro; IPR00103; -  
DR InterPro; IPR001100; -  
DR InterPro; IPR001327; -  
DR Pfam; PF00070; pyr\_redox; 1.  
DR PRINTS; PR00368; FADPNR.  
DR PRINTS; PR00411; PNDRTASEI.  
DR PRINTS; PR00459; PNDRTASEII.  
DR PROSITE; PS00076; PYRIDINE\_REDOX\_1; 1.  
KW FAD; Flavoprotein; Oxidoreductase; Redox-active center.  
FT NON\_TER 1  
SQ SEQUENCE 565 AA; 60695 MW; B26113AE09A121DE CRC64;

Query Match 41.6%; Score 44.5; DB 10; Length 565;  
Best Local Similarity 58.8%; Pred. No. 53;  
Matches 10; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 1 GETFQVEVPSQH-IDS 16  
| | | | | : ||| |||  
Db 233 GRFTPEIPGSEHAIDS 249

RESULT 19  
O49876 PRELIMINARY; PRT; 293 AA.  
ID O49876  
AC O49876;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE CLASS III CHITINASE PRECURSOR (EC 3.2.1.14) (CHITODEXTRINASE) (1,4-BETA-POLY-N-ACETYLGLUCOSAMINIDASE) (POLY-BETA-GLUCOSAMINIDASE).  
GN CHIBI OR LUPAL.  
OS Lupinus albus (White lupine).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
OC Fabales; Fabaceae; Papilionoideae; Lupinus.  
OX NCBI\_TaxID=3870;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. ULTRA;  
RA Regalado A.P., Vidal S., Neves A., Ricardo C.P.P.,  
RA Rodrigues-Pousada C.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.  
CC -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL

```

CC EMBL; Y16415; CAA76203.1; -.
DR HSSP; P23472; 2HVM.
DR Mendeil; 28297; Lupal; Chib1; 28297.
DR InterPro; IPR001579; -.
DR Pfam; PF00192; Chitinase_2; 1.
DR PROSITE; PS01095; CHITINASE_18; 1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 293 POTENTIAL.
SQ SEQUENCE 293 AA; 31128 MW; DB1B21728F657E2F CRC64;

Query Match 41.1%; Score 44; DB 10; Length 293;
Best Local Similarity 38.1%; Pred. No. 32;
Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 GETFOVEPGSQHIDSOKKAI 21
| : | : | : | : | :
Db 142 GIDFIEAGGAHQHYDELARAL 162

RESULT 20
Q9QZL7 PRELIMINARY; PRT; 330 AA.
AC Q9QZL7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE IRF6 (FRAGMENT).
GN IRF6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=129/J;
RA Sands A.; Mak T.W.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177668; AAF0915.1; -.
DR HSSP; P23906; 2IRF.
DR InterPro; IPR001346; -.
DR Pfam; PF00605; IRF; 1.
DR PRINTS; PR00267; INTERNRGFGCT.
DR PROSITE; PS00601; IRF; 1.
DR SMART; SM00348; IRF; 1.
FT NON_TER 1
FT NON_TER 330
SQ SEQUENCE 330 AA; 37355 MW; EB02EC8B751CBD7D CRC64;

Query Match 41.1%; Score 44; DB 11; Length 330;
Best Local Similarity 41.2%; Pred. No. 36;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 5 QVEVPGSQHIDSOKKAI 21
| : | : | : | : | :
Db 250 QVKEPGPEHITNEKQKL 266

RESULT 21
Q9U3M9 PRELIMINARY; PRT; 374 AA.
AC Q9U3M9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE C40H5.3 PROTEIN.
GN C40H5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Feloderinae; Caenorhabditis.

CC EMBL; Y16415; CAA76203.1; -.
DR HSSP; P23472; 2HVM.
DR Mendeil; 28297; Lupal; Chib1; 28297.
DR InterPro; IPR001579; -.
DR Pfam; PF00192; Chitinase_2; 1.
DR PROSITE; PS01095; CHITINASE_18; 1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 293 POTENTIAL.
SQ SEQUENCE 293 AA; 31128 MW; DB1B21728F657E2F CRC64;

Query Match 41.1%; Score 44; DB 10; Length 293;
Best Local Similarity 38.1%; Pred. No. 32;
Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 GETFOVEPGSQHIDSOKKAI 21
| : | : | : | : | :
Db 142 GIDFIEAGGAHQHYDELARAL 162

RESULT 20
Q9QZL7 PRELIMINARY; PRT; 330 AA.
AC Q9QZL7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE IRF6 (FRAGMENT).
GN IRF6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=129/J;
RA Sands A.; Mak T.W.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177668; AAF0915.1; -.
DR HSSP; P23906; 2IRF.
DR InterPro; IPR001346; -.
DR Pfam; PF00605; IRF; 1.
DR PRINTS; PR00267; INTERNRGFGCT.
DR PROSITE; PS00601; IRF; 1.
DR SMART; SM00348; IRF; 1.
FT NON_TER 1
FT NON_TER 330
SQ SEQUENCE 330 AA; 37355 MW; EB02EC8B751CBD7D CRC64;

Query Match 41.1%; Score 44; DB 11; Length 330;
Best Local Similarity 41.2%; Pred. No. 36;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 5 QVEVPGSQHIDSOKKAI 21
| : | : | : | : | :
Db 250 QVKEPGPEHITNEKQKL 266

RESULT 21
Q9U3M9 PRELIMINARY; PRT; 374 AA.
AC Q9U3M9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE C40H5.3 PROTEIN.
GN C40H5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Feloderinae; Caenorhabditis.

```

```

OX NCBI_TaxID=6239;
RN 1;
RP SEQUENCE FROM N.A.
RA White S.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN 12;
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81482; CAB03954.2; -.
SQ SEQUENCE 374 AA; 42198 MW; 67D202886D6A7824 CRC64;

Query Match 41.1%; Score 44; DB 5; Length 374;
Best Local Similarity 42.1%; Pred. No. 41;
Matches 8; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 GETFOVEPGSQHIDSOKK 19
| : | : | : | : | :
Db 230 GVEFHIWFGPHADEQKE 248

RESULT 22
Q9N136 PRELIMINARY; PRT; 467 AA.
AC Q9N136;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE INTERFERON REGULATORY FACTOR 6.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN 1;
RP SEQUENCE FROM N.A.
RC TISSUE=15 DAY PREGNANT UTERUS;
RA Choi Y., Spencer T.E., Bazer F.W.;
RL "Cloning and Analysis of Ovine IRF-6";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF228446; AAF34782.1; -.
DR InterPro; IPR001346; -.
DR Pfam; PF00605; IRF; 1.
DR PRINTS; PR00267; INTERNRGFGCT.
DR PRODOM; PD002355; -.
DR PROSITE; PS00601; IRF; 1.
DR SMART; SM00348; IRF; 1.
SQ SEQUENCE 467 AA; 52970 MW; 21E04F749844D88F CRC64;

Query Match 41.1%; Score 44; DB 6; Length 467;
Best Local Similarity 41.2%; Pred. No. 53;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 5 QVEVPGSQHIDSOKKAI 21
| : | : | : | : | :
Db 273 QVKEPGPEHITNEKQKL 289

RESULT 23
P97431 PRELIMINARY; PRT; 467 AA.
ID P97431;
AC P97431;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE INTERFERON REGULATORY FACTOR 6.
GN MIRF6.
OS Mus musculus (Mouse).

```





**This Page Blank (uspto)**



78511  
**STIC-Biotech/ChemLib**

**From:** Ford, Vanessa  
**Sent:** Wednesday, October 23, 2002 12:15 PM  
**To:** STIC-Biotech/ChemLib  
**Subject:** Sequence search 09/786648

Please search SEQ ID Nos: 2, 3, 4, and 5.

Please include interference searches.

Vanessa L. Ford  
Biotechnology Patent Examiner  
Office: CM1 8B13  
Mailbox: CM1 8E12  
Phone: 703.308.4735

CRFF

Point of Contact:  
Susan Hanley  
Technical Info. Specialist  
CM1 6B05 Tel: 305-4053

**This Page Blank (uspto)**